



UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

Corso di laurea in Scienze Statistiche, Economiche, Finanziarie ed Aziendali

TESI DI LAUREA

RICONCILIAZIONE DI SISTEMI DI SERIE STORICHE ECONOMICHE

Reconciliation of systems of economic time series

Relatore

Prof. Tommaso Di Fonzo

Laureando

Nicola Sturaro Sommacal

ANNO ACCADEMICO 2008/2009

Per informazioni, chiarimenti o per segnalare errori potete contattare l'autore: Nicola Sturaro Sommacal (nicola@sturaro.net)

Indice

Premessa	5
Introduzione alla tesi	5
Per i non addetti ai lavori.....	6
1 Il problema	8
1.1 Il sistema dei conti nazionali	8
1.2 La matrice di contabilità sociale	9
1.3 La tavola Input-Output e la Matrice di Contabilità Nazionale	11
1.4 Bilanciamento, benchmarking e riconciliazione	12
2 Il bilanciamento di matrici	14
2.1 Il bilanciamento di matrici.....	14
2.2 Il problema del bilanciamento	14
2.3 La procedura RAS	15
2.3.1 Limiti ed evoluzioni del RAS.....	16
2.3.2 L'aggiustamento proporzionale <i>Plus-Minus</i>	17
2.4 La procedura di Stone	18
2.4.1 Alcune note sulla matrice G	20
2.4.2 L'interpretazione statistica	21
2.4.3 Scelta della matrice V	23
2.4.4 Problemi che possono emergere con il metodo di Stone	23
2.4.5 La procedura di Stone come approssimazione del metodo RAS	25
2.5 La procedura DSS.....	26
3 Il benchmarking di serie storiche.....	27
3.1 L'approccio basato sulla minimizzazione quadratica	27

3.2	Scelta della matrice M	29
3.3	La variante proporzionale.....	32
3.4	Riferimento al saggio di variazione nel principio di conservazione del movimento	33
3.5	Conclusioni sul principio di conservazione del movimento di Denton	35
4	Riconciliazione di un sistema di serie storiche.....	37
4.1	Il problema della riconciliazione.....	37
4.2	Vincoli contemporanei	40
4.3	La stima della serie	41
4.4	Scelta della matrice V	43
4.5	Un approccio a due passi.....	44
4.6	Riconciliazione: un approccio basato sulla regressione	47
5	Un'applicazione: l'indagine sul commercio al dettaglio	50
5.1	Il Canada	50
5.2	L'indagine mensile sul commercio al dettaglio in Canada	53
5.2.1	Il campionamento	54
5.2.2	I questionari	54
5.2.3	Mancate risposte	55
5.2.4	Raccolta dei dati.....	56
5.3	La procedura di destagionalizzazione.....	57
5.4	La struttura dei dati analizzati	58
5.5	La riconciliazione del sistema MRTS per provincia/territorio con un vincolo esogeno	68
5.5.1	Vincoli temporali e benchmark univariato	68
5.5.2	Vincolo di aggregazione contemporanea esogeno	70
5.5.3	Procedure di riconciliazione e valutazione dei risultati.....	72

5.5.4	Esito e confronto dei risultati	74
5.6	La riconciliazione del sistema per settore commerciale con un vincolo esogeno	79
5.7	La riconciliazione in presenza di discrepanze temporali	80
5.8	La riconciliazione dei sistemi per provincia/territorio e per settore commerciale con un vincolo endogeno	82
5.8.1	La scelta del vincolo di aggregazione contemporanea	82
5.8.2	Valutazione delle performance	83
6	Conclusione.....	86
Appendice A.	Presentazione dei dati	91
A.1	Struttura di una matrice di contabilità sociale	91
A.2	Struttura dei dati	92
Appendice B.	Approfondimenti	93
B.1	Utilizzo dell'aggiustamento proporzionato Plus-Minus	93
B.2	Algoritmo per lo scaling DSS.....	94
B.3	Il vincolo di aggregazione temporale e dati contraddittori	95
B.4	Uso dei numeri pseudo casuali nella generazione di vincoli temporali.....	96
B.5	La funzione obiettivo e le matrici di aggregazione in presenza di serie di lunghezza diversa.	97
Appendice C.	Alcuni script di esempio.....	100
C.1	Bilanciamento di una matrice con il metodo Stone	100
C.2	Aggiustamento secondo la procedura univariata di Denton	104
C.3	Il modello di regressione per la riconciliazione	106
C.4	Funzione Matlab [®] per il benchmarking univariato secondo la procedura di Denton	109
Bibliografia.....		113

Premessa

Introduzione alla tesi

La maggior parte dei dati ottenuti dai produttori ufficiali di statistiche (ad esempio, banche centrali ed istituti di statistica) devono essere aggiustati, corretti o comunque elaborati per arrivare a valori affidabili, utili e pubblicabili.

Un problema che spesso si presenta alle agenzie statistiche riguarda il bilanciamento di tavole di serie storiche: gli istituti raccolgono e pubblicano serie mensili (o trimestrali) che devono essere conformi alle serie annuali e, contemporaneamente, soddisfare vincoli espressi da relazioni contabili. Da un punto di vista formale si è interessati a stimare i valori assunti dalle singole variabili, nel rispetto dei vincoli di aggregazione temporali e contemporanei (chiamati anche vincoli cross-sezionali).

Nel primo capitolo viene presentato, in maniera descrittiva, il problema. I primi paragrafi descrivono le tavole di contabilità nazionale, con particolare attenzione al sistema dei conti nazionali, alla Matrice di Contabilità Sociale (SAM), alla tavola Input-Output e alla Matrice di Contabilità Nazionale (NAM). Successivamente vengono introdotti i concetti di bilanciamento, benchmarking e riconciliazione.

Il bilanciamento di matrici (capitolo 2) consiste nell'aggiustare i valori dei singoli elementi affinché la loro somma per riga e per colonna coincida con i totali marginali, rispettivamente, di riga e di colonna. Gli algoritmi che vengono presentati sono due: l'algoritmo di scaling RAS e il metodo di riconciliazione ai minimi quadrati introdotto da Stone et al. (1942).

Il benchmarking univariato di serie storiche (capitolo 3) si applica per aggiustare le somme annuali di dati ad alta frequenza (tipicamente mensili o trimestrali) con i relativi benchmark annuali. La procedura, sviluppata per serie univariate da Denton (1971), è stata successivamente ripresa da diversi autori.

Se gli argomenti dei due capitoli precedenti sono ormai consolidati nel loro uso, i problemi emergono quando i due problemi si presentano contemporaneamente: si è interessati a bilanciare una tavola, in cui una delle due dimensioni rappresenta il tempo. È il problema della riconciliazione di sistemi di serie storiche economiche (capitolo 4), argomento centrale della tesi. Le procedure proposte sono due: l'estensione multivariata della procedura di Denton, dovuta a Di Fonzo, che estende al caso multivariato la metodologia vista nel capitolo 3 e la procedura dovuta a Quenneville e Rancourt (2005), che prevede prima l'applicazione univariata (serie per serie) della metodologia di Denton e quindi il bilanciamento con la procedura di Stone et al. (1942).

Il capitolo 5 propone un'applicazione delle metodologie di riconciliazione. Ripetiamo, con le due procedure viste al capitolo precedente, il lavoro di riconciliazione effettuato dagli esperti di Statistics Canada per aggiustare le tavole dell'indagine mensile sul commercio al dettaglio (MRTS).

Per i non addetti ai lavori

I dati raccolti, ad esempio, dai produttori ufficiali di statistiche (ISTAT, Banca d'Italia ecc.) sono spesso rappresentati sotto forma di tavole, in cui le righe rappresentano una certa classificazione e le colonne un'altra.

Si pensi ad esempio ai dati relativi alle vendite al dettaglio di una certa nazione (questo esempio, con riferimento al Canada, verrà proposto nel capitolo 5). In questo caso le righe possono rappresentare i settori commerciali (supermercati, venditori di auto, stazioni di servizio e così via) e le colonne diverse aree geografiche (province o regioni, ad esempio). I *totali marginali*, cioè i totali di riga e di colonna, rappresentano, in questo caso, il totale di ciascun settore commerciale o di ciascuna area geografica. Il totale complessivo, rappresenta il totale delle vendite al dettaglio della nazione considerata. Se questi dati provengono da fonti diverse, o sono stati sottoposti a trattamenti statistici (la *destagionalizzazione*, ad esempio), è possibile che non soddisfino i *vincoli*; cioè che i totali di riga o di colonna e/o il totale complessivo non siano ottenibili come somma delle diverse righe e/o colonne. Gli uten-

ti, invece, preferiscono disporre di dati in cui vengono soddisfatte le relazioni naturali: è allora necessario aggiustare i dati, perché ciò accada. Le procedure di questo tipo, dette procedure di *bilanciamento*, sono illustrate nel capitolo 2.

Si pensi adesso, invece, a un singolo fenomeno di interesse, di cui si hanno sia dati ad alta frequenza (mensili o trimestrali) che dati annuali, raccolti spesso con metodologie diverse. Anche in questo caso è necessario aggiustare i dati mensili (o trimestrali) affinché la loro somma coincida col dato annuale, senza però alterare, o facendolo il meno possibile, la dinamica temporale. Questa operazione, che chiameremo *benchmarking*, è presentata nel capitolo 3.

Il caso concettualmente più complicato è quello di tavole, come nel primo esempio, in cui le righe rappresentano il tempo e le colonne un certo criterio di classificazione (ad esempio un'area geografica o un settore commerciale). In questo caso si devono aggiustare i dati affinché i totali di colonna coincidano con la somma dei valori interni e i totali annuali coincidano con la somma dei dati mensili (o trimestrali), senza stravolgere la dinamica temporale. Questa operazione è detta di *riconciliazione di sistemi di serie storiche economiche*, e dà il titolo a questa tesi. È illustrata nel capitolo 4, mentre il capitolo 5 propone un'applicazione delle diverse tecniche illustrate a dati reali.

1 Il problema

I problemi di bilanciamento emergono quando dei dati non soddisfano dei vincoli, che spesso derivano dalla natura dei dati stessi. Questi problemi sono frequenti in statistica economica, e più precisamente nella contabilità nazionale, dove si deve lavorare con tavole, spesso di dimensioni notevoli, che devono soddisfare una serie di vincoli contabili. Altre applicazioni si hanno in svariate altre discipline, tra cui nel campo dei trasporti (stima di matrici partenze-arrivi), della statistica (tabelle di contingenza), della demografia (flussi migratori) e del calcolo delle probabilità (stima delle probabilità di transizione).

1.1 Il sistema dei conti nazionali

L'Istat (2007a) definisce così il sistema dei conti nazionali: *«Il sistema dei conti nazionali descrive in termini quantitativi e sotto forma contabile l'attività economica e finanziaria di un paese o di sue ripartizioni territoriali, per periodi coincidenti con l'anno solare e per intervalli temporali più brevi».*

Alcuni eventi fondamentali del dopoguerra, e in particolare l'intensificazione delle relazioni internazionali, hanno fatto sorgere l'esigenza di una standardizzazione delle modalità di raccolta, codifica e presentazione dei dati. Per soddisfare queste esigenze, nel 1953, l'ONU pubblicò il *System of National Accounts* (SNA) per i paesi ad economia di mercato ed il *Material Product System* (MPS) per i paesi ad economia collettivista. Lo SNA fu in seguito modificato, accogliendo il suggerimento di Stone di presentare i dati in forma matriciale, nel 1968.

Introdotta, nell'accezione attuale, da Stone (1962) per l'economia britannica nell'ambito del *Cambridge Growth Project*, la matrice di contabilità sociale (SAM, *Social Accounting Matrix*) all'inizio degli anni '70 si caratterizza per una posizione antagonista rispetto allo SNA, con il quale però condivide concetti, definizioni e classificazioni; la SAM è infatti uno strumento flessibile che poteva essere adattato

alle diverse realtà locali, quindi non in linea con il principio di standardizzazione che era alla base dello SNA.

Nel 1970 l'Istituto Statistico delle Comunità Europee (oggi Eurostat) elaborò il Sistema Europeo di Conti nazionali e regionali (SEC). Il SEC è la trasposizione, più consona per l'utilizzo in seno all'Unione Europea, dello SNA, dal quale differisce per una più puntuale definizione di alcune regole, utile per ridurre i problemi di interpretazione ed assicurare un livello di comparabilità ancora più elevato alle statistiche dell'Unione Europea.

Nel corso del primo ventennio di applicazione sia SNA sia SEC hanno dimostrato di avere parti superate dagli eventi storici (quali, ad esempio, l'alta inflazione e lo sfruttamento intensivo di alcune risorse) e parti di controversa interpretazione o addirittura insoddisfacenti. Pertanto è stato redatto dalle massime autorità a livello mondiale in materia di statistica (l'ONU, il Fondo Monetario Internazionale, la Commissione delle Comunità Europee -tramite Eurostat-, l'OCSE e la Banca Mondiale) un'ulteriore modifica dello SNA, lo SNA93, che ha portato nel 1995 alla versione attuale del SEC.

Nei prossimi paragrafi sarà analizzata più nel dettaglio la matrice di contabilità sociale (SAM) e altri importanti strumenti della contabilità nazionale, la tavola Input-Output e la matrice di contabilità nazionale: è possibile vedere la SAM come il completamento della prima e l'evoluzione della seconda.

1.2 La matrice di contabilità sociale

La matrice di contabilità sociale è una tavola a doppia entrata¹ che registra i flussi che intercorrono tra gli operatori di un sistema economico. Fu introdotta da Stone nel 1960: prima di questa data, informazioni dettagliate sul sistema economico erano concentrate nei conti della produzione ed assumevano la forma di tavole input-output, anche se c'erano già state alcune esperienze simili relative alla Norvegia e ai Paesi Bassi, rispettivamente, degli anni '30 e '40 (Pyatt e Round, 1977).

¹ La struttura della tavola è riportata nell'appendice A.1.

Nel quadro delle revisioni dei sistemi contabili, la SAM ha assunto sempre più una posizione di rilievo, anche nella stesura dei conti nazionali. Può essere considerata, infatti, un vero e proprio mezzo per creare la Contabilità Nazionale, perché i suoi conti sono compilati utilizzando fonti diverse e alla costruzione segue una procedura di bilanciamento che consente di pervenire ad un complesso sistematico e coerente di informazioni.

Malgrado la SAM nasca come strumento per i paesi in via di sviluppo, negli anni più recenti è divenuta oggetto di interesse anche per le economie sviluppate, perché fornisce un quadro contabile completo di un sistema economico e consente di costruire modelli di equilibrio economico generale partendo dalla matrice di contabilità stessa. Altro elemento che la rende uno strumento estremamente attuale è l'enfasi posta sulla distribuzione del reddito. Questa è al contempo causa ed effetto: causa del consumo e del risparmio, cioè dei comportamenti individuali, ed effetto, perché a sua volta influenzata dalle modifiche di variabili esogene (un aumento delle esportazioni influenzerà, ad esempio, la distribuzione del reddito) ed in particolare di alcune delle variabili strumentali (tutte le azioni di politica economica hanno anche un effetto redistributivo). Nessuna fonte è in grado di fornire sufficienti informazioni su tutti i fenomeni di un sistema economico, quindi, la costruzione della SAM -che descrive appunto un sistema economico- richiede l'utilizzo di diverse fonti. Questo comporta la necessità di integrare dati di natura e provenienza diversa. L'integrazione è il momento fondamentale nel processo di costruzione di una matrice di contabilità sociale, e si caratterizza proprio per la capacità di valorizzare le statistiche relative a fenomeni specifici solitamente indipendenti. La SAM può essere considerata un vero e proprio mezzo per creare la Contabilità Nazionale. La flessibilità della SAM riguarda, sostanzialmente, la possibilità di disaggregare gli intestatari dei conti ricorrendo, caso per caso, ai criteri classificatori più idonei; è possibile decidere quali aspetti privilegiare e con che grado di dettaglio, anche in relazione ad un eventuale modello da utilizzare per la simulazione di politiche economiche e sociali. Ha il pregio di comprendere tutte le fasi del processo economico e di evidenziarne la circolarità; non si limita, inoltre, alle sole transazioni interne ad un paese, ma comprende i flussi da e verso il resto del mondo.

Costruire una SAM richiede l'utilizzo di una pluralità di fonti di diversa natura; le principali diffuse dall'ISTAT sono i Conti Economici Nazionali Annuali (CENA), i Conti dei Settori Istituzionali e la Tavola Intersectoriale dell'Economia Italiana (TEI) che costituiscono, in linea generale, i vincoli da rispettare. Il limite di queste fonti è di fornire dati solitamente aggregati per i vari fenomeni e non consentire, quindi, di entrare nella specificità di tematiche quali la distribuzione, il consumo ecc. Da qui il ricorso ad indagini specifiche come quella della Banca d'Italia sul reddito oppure quella Istat sul consumo.

1.3 La tavola Input-Output e la Matrice di Contabilità Nazionale

Altri strumenti, oltre alla SAM, descrivono le transazioni che intercorrono in un sistema economico, o parte di esse; questi sono la tavola Input-Output e la Matrice di Contabilità Nazionale (NAM).

La tavola Input-Output (o delle interdipendenze produttive) nasce nel 1936 ad opera di Leontief. Ha una forma matriciale a doppia entrata, scomponibile in tre sotto-matrici. La prima, incrocio branche-branche, costituisce l'elemento caratterizzante la tavola stessa; la seconda sub-matrice, incrocio branche-fattori produttivi, registra i redditi primari che remunerano i fattori produttivi impiegati nella produzione. La terza sub-matrice registra flussi finali, ossia l'incrocio tra branche e domanda finale (consumo privato, spesa pubblica, investimenti ed esportazioni). La tavola Input-Output non comprende il processo distributivo e redistributivo del reddito, aspetto chiave, invece, nella SAM grazie all'inclusione dei conti intestati ai settori istituzionali. La SAM completa, quindi, la tavola Input-Output aggiungendo le matrici che consentono di chiudere il modello di Leontief, tenendo presente le relazioni tra distribuzione funzionale e personale del reddito, e composizione della spesa per consumo. I conti intestati ai settori istituzionali consentono di cogliere, appunto, oltre all'intero processo di formazione del reddito disponibile, l'atteggiamento degli stessi verso consumo ed investimenti.

Lo SNA93 prevede la rappresentazione dei conti nazionali in forma matriciale (National Accounting Matrix, NAM). La NAM ha una struttura rigida, perché fa riferi-

mento ad una pura sequenza di conti e di tipologie di unità intestatarie degli stessi, come previsto in contabilità; non privilegia, inoltre, aspetti particolari. La SAM si differenzia dalla NAM per la maggiore flessibilità che va ricercata, oltre che nella possibilità di scegliere quali fasi del processo economico evidenziare, nell'opportunità di disaggregare gli intestatari dei conti ricorrendo ai criteri classificatori più idonei distinguendo caso per caso; la SAM pur essendo una sequenza di conti, consente di enfatizzare formazione, distribuzione e redistribuzione del reddito.

1.4 Bilanciamento, benchmarking e riconciliazione

Si sono finora visti i più diffusi strumenti impiegati nella contabilità nazionale. Questi strumenti sono utilizzati non solo con riferimento a singoli istanti temporali, ma sotto forma di sistemi di serie storiche.

Le statistiche economiche sono spesso collegate da un sistema di relazioni contabili. Alcune restrizioni lineari sono originate dalla teoria economica (per esempio, il prodotto interno lordo come saldo tra impieghi e risorse), altre sono dovute al livello di disaggregazione col quale sono fornite le statistiche (per esempio il valore aggiunto per l'economia totale, cioè per il complesso di unità o settori istituzionali, è la somma dei valori aggiunti delle 21 macrobranche di attività economica previste dalla classificazione NACE). In ogni caso, questi vincoli sono raramente soddisfatti dai dati raccolti: questo accade perché, ad esempio, i dati economici sono raccolti con metodi diversi, da campioni diversi, elaborati con tecniche diverse ed infine pubblicati per scopi diversi. Quindi i valori ottenuti presentano delle discrepanze rispetto ai vincoli che dovrebbero soddisfare; tali discrepanze creano confusione tra gli utenti e imbarazzo ai produttori dei dati.

Analogamente, gli istituti sono spesso obbligati a pubblicare set di serie storiche consistenti per soddisfare obblighi legali o pratiche comuni nei confronti delle istituzioni internazionali (Nazioni Unite, FMI, Eurostat, ecc.).

Le restrizioni contabili possono essere di due tipi: i vincoli contemporanei, secondo cui un aggregato è la somma delle altre variabili in ogni istante temporale, e i vincoli

di aggregazione temporale, necessari quando le serie preliminari ad alta frequenza non sono in linea con i valori degli aggregati a bassa frequenza considerati noti².

La struttura, la più generale, dei dati può quindi essere riassunta nel modo seguente:

- m serie storiche di dati p ad alta frequenza (tipicamente trimestrali);
- m serie storiche di dati z_t a bassa frequenza (tipicamente annuali), che rappresentano i vincoli di aggregazione temporale;
- H serie storiche z_c ad alta frequenza, che rappresentano i vincoli di aggregazione contemporanei.

La tabella a pagina 92 aiuta a capire il problema.

Le procedure statistiche per ripristinare la consistenza tra le variabili (*between*) e entro le variabili (*within*) sono molto simili, ma le due problematiche sono spesso trattate separatamente in letteratura. In particolare, il primo problema è noto come *balancing problem*, mentre il processo di aggiustamento nella dimensione temporale è detto *benchmarking* (o disaggregazione temporale) di serie storiche.

L'aggiustamento di un set di dati per soddisfare entrambi i tipi di vincolo è conosciuto come *reconciliation problem*, che si può tradurre come "problema di riconciliazione". Il problema principale in questo caso è dovuto essenzialmente alla dimensione delle matrici coinvolte: infatti, all'aumentare del numero di variabili, della lunghezza delle serie o del numero di vincoli, il tempo di calcolo può diventare molto elevato.

L'aggiornamento e il bilanciamento di matrici di input-output o dei conti sociali ed economici o il *benchmarking* delle serie trimestrali dei conti nazionali (QNA, *Quarterly National Accounts*) agli aggregati annuali sono, al giorno d'oggi, problemi affrontati comunemente dalle agenzie statistiche. Più in generale, molte statistiche economiche pubblicate hanno superato un processo di riconciliazione.

² Ovviamente anche questi valori sono frutto di rilevazioni statistiche e pertanto soggetti ad errori. Questi dati sono però raccolti con metodi più uniformi e con maggiori obblighi legali dai vari istituti nazionali di statistica e sono quindi maggiormente affidabili.

2 Il bilanciamento di matrici

2.1 Il bilanciamento di matrici

Sia P una matrice $(n \times m)$,

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nm} \end{bmatrix},$$

e siano $z_c = [z_{c1} z_{c2} \cdots z_{cm}]'$ e $z_r = [z_{r1} z_{r2} \cdots z_{rn}]'$ due vettori, $(m \times 1)$ e $(n \times 1)$ rispettivamente.

I problemi che si possono presentare più frequentemente sono:

- (1) Trovare una matrice Y , di dimensione $(n \times m)$, tale che

$$\sum_{j=1}^m y_{ij} = z_{ri} \quad \text{per } i = 1, \dots, n$$

$$\sum_{i=1}^n y_{ij} = z_{cj} \quad \text{per } j = 1, \dots, m$$

- (2) Posto, $m = n$, trovare una matrice Y tale che

$$\sum_{j=1}^n y_{kj} = \sum_{i=1}^n y_{ik} \quad \text{per } k = 1, \dots, n$$

2.2 Il problema del bilanciamento

Nel primo capitolo si è discusso del problema, ora si inizia a parlare di possibili soluzioni. Possiamo riassumere il problema in: «data una matrice rettangolare P determinare una matrice Y che sia “il più vicino” possibile a P e soddisfi un insieme di re-

strizioni lineari sui suoi valori»¹. Il problema è noto come bilanciamento (dall'inglese *balancing*) della matrice P . Una matrice Y si dice bilanciata se soddisfa le restrizioni lineari che caratterizzano il problema.

Gli algoritmi per il bilanciamento possono essere separati in due categorie:

- gli algoritmi di *scaling*, come il RAS e il DSS, moltiplicano la matrice per una costante positiva fino a bilanciarla,
- gli algoritmi di ottimizzazione, minimizzano una funzione di perdita (o funzione criterio) che misura la distanza tra una matrice candidata ad essere la matrice bilanciata e la matrice originaria. Le condizioni del bilanciamento sono vincoli del modello di ottimizzazione: in questo modo la matrice che si ottiene è la matrice bilanciata “più vicina” a P , essendo il concetto di vicinanza quello introdotto dalla metrica della funzione di perdita scelta.

2.3 La procedura RAS

La procedura RAS è un algoritmo di *scaling* per risolvere il primo dei due problemi presentati, nel quale una matrice P non negativa ($n \times m$) deve essere aggiustata affinché i totali di riga e di colonna siano uguali a due vettori positivi, z_r e z_c , considerati noti.

Il metodo RAS è stato sviluppato ed utilizzato in maniera indipendente da numerosi ricercatori, in diverse nazioni e per diversi problemi: Kruithof (1937) propose l'algoritmo col nome di *method of twin factors* in una procedura per predire il traffico telefonico; Deming e Stephan (1940) lo usarono, chiamandolo *method of iterative proportions*, per trovare una soluzione approssimata al problema dei minimi quadrati per stimare le frequenze di una tavola di contingenza della quale sono noti i totali. Si dice² che il nome derivi dalla matrice A , da aggiustare, e dalle matrici di pre e post moltiplicazione R e S , usate da Stone (1961) nei suoi appunti.

L'algoritmo funziona nel modo seguente:

¹ Tradotto da Schneider e Zenios (1990).

² Per esempio in Lahr e De Mesnard (2004), dove si fa anche notare come il nome delle matrici di pre e post moltiplicazione sia, forse non casualmente, uguale alle iniziali dell'autore, Richard Stone.

1. *Input*. Una matrice non negativa P , $(n \times m)$, e due vettori z_r e z_c di dimensioni $(n \times 1)$ e $(m \times 1)$, rispettivamente.
2. *Passo zero (Inizializzazione)*. Si ponga $k = 0$ e $Y^0 = P$.
3. *Primo passo (Scaling per riga)*. Per $i = 1, \dots, n$ si definisca

$$\rho_i^k = \frac{z_{ri}}{\sum_j y_{ij}^k}$$

e si aggiorni Y^k con $y_{ij}^k \leftarrow \rho_i^k y_{ij}^k$ per $i = 1, \dots, n$ e $j = 1, \dots, m$.

4. *Secondo passo (Scaling per colonna)*. Per $j = 1, \dots, m$ si definisca

$$\sigma_j^k = \frac{z_{cj}}{\sum_i y_{ij}^k}$$

e si definisca Y^{k+1} con $y_{ij}^{k+1} \leftarrow \sigma_j^k y_{ij}^k$ per $i = 1, \dots, n$ e $j = 1, \dots, m$.

5. *Terzo passo*. Si sostituisca $k \leftarrow k + 1$ e si riparta dal primo passo.

Bregman (1967) ha dimostrato che, se $p_{ij} > 0$, la sequenza $\{Y^k\}$ generata dall'algoritmo converge all'unica matrice bilanciata che minimizza

$$f(y, p) = \sum_{j=1}^m \sum_{i=1}^n y_{ij} \ln \left(\frac{y_{ij}}{p_{ij}} \right) \quad (2.1)$$

rispettando i vincoli

$$\sum_{j=1}^m y_{ij} = z_{ri} \quad \text{per } i = 1, \dots, n \quad (\text{totali di riga})$$

$$\sum_{i=1}^n y_{ij} = z_{cj} \quad \text{per } j = 1, \dots, m \quad (\text{totali di colonna}).$$

2.3.1 Limiti ed evoluzioni del RAS

Quando è possibile applicarlo il metodo RAS ha il vantaggio di essere semplice da utilizzare, ma questa semplicità comporta numerosi svantaggi. Tra le critiche più frequentemente mosse al RAS si sottolineano l'assenza di significato economico (ma a questo proposito si veda il paragrafo 2.4.5) e l'impossibilità:

- (1) di bilanciare a qualcosa di diverso dai totali di riga e colonna,

- (2) di utilizzare una qualsiasi combinazione lineare delle variabili, ma solo la loro somma,
- (3) di poter considerare alcuni dei valori preliminari più affidabili degli altri;
- (4) di poter essere applicato direttamente a valori negativi.

Si noti che il quarto aspetto deriva dal secondo: se fosse possibile applicare il RAS a una combinazione lineare delle variabili allora lo si potrebbe applicare anche a valori negativi.

Per superare questo grosso limite sono state proposte numerose alternative basate sul metodo RAS. La più usata consiste nel non considerare i valori negativi nell'applicazione del RAS, e di reintrodurli nella matrice bilanciata. Questa procedura non minimizza la perdita di informazione di Kullback. Successivamente sono state proposte altre procedure: il GRAS di Junius e Oosterhaven (2003) e le soluzioni basate sulla minimizzazione delle differenze assolute e quadratiche di Jackson e Murray (2004).

Oosterhaven e Talsma (2005) hanno comparato empiricamente queste alternative concludendo che la soluzione migliore resta quella di considerare i valori negativi come esogeni e reinserirli al termine del bilanciamento. Un'altra possibilità (Di Fonzo e Marini, 2007a) consiste nell'adattare al RAS la tecnica di aggiustamento proporzionale *Plus-Minus*.

2.3.2 L'aggiustamento proporzionale *Plus-Minus*

La tecnica di *Plus-Minus proportionate adjustment* (PMPA) è stata introdotta dall'Australian Bureau of Statistics³, ed è qui riportata nella versione adattata da Di Fonzo e Marini (2007a).

Il PMPA è una procedura per bilanciare una sequenza di valori positivi e negativi a un certo totale che minimizza lo scostamento rispetto ai valori originari. Si basa sull'uso di coefficienti di aggiustamento:

³ Australian Bureau of Statistics (1995). Demographic estimates and projections: concepts, sources and methods.

$$\alpha_i = 1 + \operatorname{sgn}(p_i) \frac{(z - \sum_{i=1}^n p_i)}{\sum_{i=1}^n |p_i|}.$$

Una volta calcolati i coefficienti è sufficiente moltiplicarli per ciascun valore. Un esempio è riportato nell'appendice B.1. Osservando la formula, e soprattutto l'esempio, ci si rende conto di come i coefficienti non siano diversi per ciascun valore, ma siano soltanto due: il coefficiente di aggiustamento per valori positivi

$$\alpha^+ = 1 + \frac{(z - \sum_{i=1}^n p_i)}{\sum_{i=1}^n |p_i|}$$

e il coefficiente di aggiustamento per valori negativi

$$\alpha^- = 1 - \frac{(z - \sum_{i=1}^n p_i)}{\sum_{i=1}^n |p_i|}.$$

2.4 La procedura di Stone

Il metodo di ottimizzazione ai minimi quadrati è stato introdotto da Stone (1961) e successivamente ripreso da Stone stesso e da Byron (1978). Si basa sulla minimizzazione di una funzione criterio soggetta al vincolo che esprime il problema di minimizzazione. La funzione obiettivo di questi problemi di minimizzazione può assumere diverse forme. Si considerino ora questi due semplici criteri:

- (1) quello basato sulla minimizzazione del quadrato della distanza euclidea tra Y (la matrice che si vuole ottenere) e P (la matrice preliminare), proposto da Almon (1968), e
- (2) quello basato sul χ^2 di Pearson, utilizzato da Deming e Stephan (1940) e Friedlander (1961).

In maniera più formale, Almon (1968) considera la funzione obiettivo

$$f(y, p) = \sum_{i=1}^n \sum_{j=1}^m (y_{ij} - p_{ij})^2,$$

mentre la quantità da minimizzare secondo Deming e Stephan (1940) e Friedlander (1961) è

$$f(y, p) = \sum_{i=1}^n \sum_{j=1}^m \frac{(y_{ij} - p_{ij})^2}{p_{ij}}.$$

Si noti come questa classe di metodi non sempre preserva il segno: può generare valori negativi anche se non ne sono presenti nella matrice preliminare. Si può tuttavia riconoscere immediatamente che, data P matrice non negativa, entrambe le *penalty function* sono funzioni da minimizzare *Quadratic Positive Definite* (QPD)⁴ della forma

$$(p - y)' Q (p - y) \quad (2.2)$$

dove p e y sono entrambi vettori ($nm \times 1$) contenenti, rispettivamente, gli elementi di P e Y ; $Q = I_{nm}$ nella formulazione di Almon e $Q = [diag(p)]^{-1}$ nella successiva.

Sia \tilde{y} un vettore ($nm \times 1$) che soddisfa il sistema di vincoli contabili lineari indipendenti

$$G\tilde{y} = z \quad (2.3)$$

dove G è una matrice ($k \times nm$), di rango $k < nm$, e z è un vettore ($k \times 1$) di costanti note.

Sia L la funzione lagrangiana

$$L = (y - p)' Q (y - p) + 2\lambda' (Gy - z)$$

dove λ è il vettore dei moltiplicatori di Lagrange e Q è una matrice simmetrica, definita positiva.

La soluzione che rende minima la funzione criterio, si veda ad esempio Luenberger (2003), è ottenuta uguagliando a 0 le derivate parziali di L rispetto a y e a λ , e risolvendo.

$$\begin{aligned} & \begin{cases} \partial L / \partial y \\ \partial L / \partial \lambda \end{cases} = \underline{0} \Rightarrow \\ \Rightarrow & \begin{cases} 2Q(y - p) + 2G'\lambda \\ 2(Gy - z) \end{cases} = \begin{cases} Qy - Qp + G'\lambda \\ Gy - z \end{cases} = \underline{0} \Rightarrow \end{aligned} \quad (2.4)$$

⁴ Harrigan (1990)

$$\Rightarrow \begin{cases} \hat{y} = Q^{-1}Qp - Q^{-1}G'\lambda = p - Q^{-1}G'\lambda \\ \hat{\lambda} = (GQ^{-1}G')^{-1}(Gp - z) \end{cases}.$$

In conclusione, si può scrivere

$$y = p + Q^{-1}G'(GQ^{-1}G')^{-1}(z - Gp). \quad (2.5)$$

Contrariamente al RAS, la procedura di Stone bilancia automaticamente quando i totali di riga e i totali di colonna sono endogeni e può essere applicata a problemi di bilanciamento dei conti nazionali e ad altre situazioni di interesse. Se consideriamo il caso di vincoli non stringenti, allora possiamo considerare z come una delle variabili p e il vincolo diventa

$$Gp = z \Rightarrow Gp - z = 0 \Rightarrow G_z p_z = 0. \quad (2.6)$$

La stima ottenuta rispetta i vincoli imposti, poiché:

$$Gy = Gp + \underbrace{GVG'(GVG')^{-1}}_{\text{matrice identità}}(z - Gp) = Gp - Gp + z = z$$

2.4.1 Alcune note sulla matrice G

L'uso della procedura di Stone per bilanciare a due vettori z_c e z_r contenenti, rispettivamente, i totali di riga e i totali di colonna, è possibile tramite una scelta opportuna della matrice G , tale che il suo prodotto per il vettore p degli elementi di P sia il vettore $z = [z_c \quad z_r]$ dei totali di colonna e di riga. Si consideri inizialmente la matrice G^* che ha la forma di una matrice a blocchi; il blocco superiore è costituito da una matrice identità di dimensioni $(m \times m)$, a sua volta a blocchi: gli elementi sulla diagonale sono vettori $(1 \times n)$ costituiti da 1, e gli elementi all'esterno della diagonale sono vettori nulli delle stesse dimensioni. Il blocco inferiore è invece costituito da un vettore $(1 \times m)$ di matrici identità $(n \times n)$. Complessivamente la matrice G^* è di dimensioni $((n + m) \times (nm))$ e ha la forma

$$G^* = \begin{bmatrix} I^* \\ \underbrace{I_{n \times n} \quad I_{n \times n} \quad \cdots \quad I_{n \times n}}_{m \text{ matrici identità}} \end{bmatrix},$$

$$I^* = \begin{bmatrix} \underline{1} & \cdots & \underline{0} \\ \vdots & \ddots & \vdots \\ \underline{0} & \cdots & \underline{1} \end{bmatrix}$$

con $\underline{1}$ e $\underline{0}$ vettori $(1 \times n)$ contenenti, rispettivamente, tutti 1 e 0.

È necessario notare a questo punto che il numero di righe della matrice G^* è pari al numero di vincoli. Si ricordi tuttavia che uno degli $n + m$ vincoli non è linearmente indipendente dagli altri: date tutte le m somme di colonna e $n - 1$ somme di riga, l'ultima somma di riga è vincolata. Per questo la matrice G da utilizzare è la matrice G^* senza l'ultima riga. Tale matrice ha dunque tante righe quanti sono i vincoli linearmente indipendenti, $(n + m - 1)$, e tante colonne quanti sono gli elementi di P , cioè nm .

Nella pratica del bilanciamento di tavole di contabilità nazionale, la matrice G è costituita non solo da elementi 0 e 1, ma anche pari a -1.

Si consideri l'esempio descritto nell'appendice C.1. In questo caso i totali di riga e di colonna vengono considerati non stringenti, quindi la matrice dei dati preliminari vettorizzata assume la forma

$$p'_z = [p'_1 \quad p'_2 \quad \cdots \quad p'_m \quad z'_r \quad z'_c],$$

dove p'_j è il j -esimo vettore colonna della matrice dei dati preliminari $P = [p_1 \quad p_2 \quad \cdots \quad p_j \quad \cdots \quad p_m]$ e z_r e z_c sono i vettori colonna dei totali di riga e di colonna, rispettivamente.

La matrice G si presenta nella forma $G = [G^* \quad -I]$, dove la matrice G^* è quella descritta ad inizio paragrafo e $-I$ è la matrice identità moltiplicata per lo scalare -1 .

2.4.2 L'interpretazione statistica

Si analizza ora la procedura di bilanciamento di Stone in un'ottica statistica. Si assume che i dati preliminari, contenuti nel vettore p , si distribuiscano correttamente attorno agli ignoti dati "veri":

$$p = \tilde{y} + e \tag{2.7}$$

dove e è un vettore ($nm \times 1$) di errori con media zero e matrice di covarianza V nota. È stato dimostrato⁵ che y ottenuto dalla (2.5) fornisce uno stimatore efficiente di \tilde{y} , che appartiene alla classe degli stimatori BLUE (*Best Linear Unbiased Estimator*), che minimizza la (2.2) per $Q^{-1} = V$ e soddisfa i vincoli (2.3):

$$y = p + VG'(GVG')^{-1}(z - Gp) \quad (2.8)$$

con

$$E(y - \tilde{y})(y - \tilde{y})' = V - VG'(GVG')^{-1}GV,$$

Poiché il secondo termine a destra è una matrice (semi) definita positiva, vi è un guadagno in attendibilità.

L'interpretazione della (2.8) è che p viene bilanciata ridistribuendo gli errori di quadratura commessi tra i suoi vari elementi secondo una funzione che assegna una porzione dell'errore proporzionale alla varianza⁶. Questa caratteristica, che differenzia questa procedura da altri metodi, per esempio il RAS, ha un'importanza pratica notevole: permette di ottenere una matrice finale Y che modifica il meno possibile gli elementi di P che godono di un'elevata affidabilità, e che effettua delle correzioni più ampie di quei dati giudicati meno attendibili. Appare chiaro quindi il ruolo determinante della matrice di covarianza nell'aggiustamento dei singoli valori in p .

Antonello (1989) dimostra lo stesso risultato in una prospettiva più interessante, perché massimizza la funzione di probabilità

$$p(y|p) \propto p(y)l(y|p),$$

nel rispetto dei vincoli (2.6) e sotto l'ipotesi di normalità. Questo permette di ottenere stime che sono di massima verosimiglianza e possono essere applicati gli strumenti standard dell'inferenza statistica.

⁵ Theil (1961)

⁶ Lavanda et al. (1999)

2.4.3 Scelta della matrice V

È assai raro che siano disponibili valutazioni sull'affidabilità delle stime preliminari o sulle covarianze fra di esse: gli elementi della matrice V sono quindi generalmente ignoti. Per questo, nella pratica, si pone un'assunzione semplificatrice sui termini d'errore, ipotizzando che essi siano indipendenti; in questo caso V è diagonale.

Un'ipotesi di lavoro di Stone, consiste nell'approssimare la varianza di ciascuna osservazione con il valore stesso, ossia

$$\begin{cases} V_{ii} = P_{ii} \\ V_{ij} = 0 \end{cases} \quad \forall i \neq j \quad (2.9)$$

Operativamente la (2.9) è modificata considerando, per ciascuna osservazione, un coefficiente di alterabilità soggettivo, indicato con a_i . Pertanto le singole varianze sono fissate pari a

$$V_{ii} = (a_i P_{ii})^2,$$

mentre le covarianze sono tutte nulle. Nella pratica corrente si preferisce limitare il range di a_i (che, teoricamente, potrebbe includere tutti i valori non negativi) all'intervallo compreso tra 0 e 1. Questo significa che, in presenza di valori giudicati poco affidabili, la varianza è pari al valore stesso della stima preliminare. Allo stesso tempo conviene prestare attenzione nell'assegnare un indicatore di inattendibilità pari a zero, per ragioni numeriche: una varianza nulla, sotto l'ipotesi di covarianze nulle, rende singolare la matrice V . Conviene pertanto assegnare ad a_i valori molto piccoli ma diversi da zero.

2.4.4 Problemi che possono emergere con il metodo di Stone

La procedura descritta presenta due principali problematiche, in parte già accennate:

- (1) non preserva il segno del valore del dato preliminare;
- (2) per strutture contabili complesse possono sorgere problemi computazionali che rendono troppo lungo, o addirittura impossibile, il calcolo diretto della (2.8).

Una possibile soluzione al secondo problema è stata proposta da Byron (1978) e, successivamente, ripresa da Nicolardi (2000) e implementata in EULERO⁷ per il bilanciamento delle matrici di contabilità nazionale dell'ISTAT. La procedura si basa sul metodo del gradiente coniugato che permette di ottenere una soluzione del sistema (2.4) che non richiede l'inversione di matrici attraverso un algoritmo iterativo. La soluzione rispetto a λ della (2.4) può essere scritta come

$$\begin{aligned} (GVG')^{-1}\lambda &= Gp - z \\ A\lambda &= b \end{aligned} \quad (2.10)$$

con $A = (GVG')^{-1}$ e $b = Gp - z$.

Ciascuna iterazione è definita da un sistema di cinque equazioni:

$$\begin{aligned} \alpha_i &= \frac{\rho'_i \rho_i}{\phi'_i A \phi_i} \\ \lambda_{i+1} &= \lambda_i + \alpha_i \phi_i \\ \rho_{i+1} &= \rho_i - \alpha_i A \phi_i \\ \beta_i &= \frac{\rho'_{i+1} \rho_{i+1}}{\rho'_i \rho_i} \\ \phi_{i+1} &= \rho_{i+1} + \beta_i \phi_i \end{aligned}$$

dove $i = 0, 1, 2, \dots$. Per l'inizializzazione dell'algoritmo si usa l'assunzione

$$\phi_0 = \rho_0 = b - A\lambda_0$$

con λ_0 contenente un vettore di condizioni iniziali per λ . La condizione di uscita dell'algoritmo si verifica controllando ad ogni iterazione la differenza $b - A\lambda$: quando questa è inferiore ad un livello di approssimazione fissato le stime λ soddisfano il sistema (2.4) e possono essere utilizzate per stimare y . Si dimostra che l'algoritmo converge al massimo in k iterazioni e che questo numero diminuisce notevolmente scalando il sistema per una opportuna matrice diagonale di pre-condizionamento.

⁷ Si tratta di un macrolinguaggio di programmazione, sviluppato come parte di un progetto finalizzato ISTAT, orientato alla definizione di una struttura a blocchi e caratterizzato da una sintassi semplice. EULERO permette di rendere l'applicazione del metodo di bilanciamento il più possibile flessibile ed immediata, definendo la struttura delle equazioni tramite un semplice file in formato testo.

2.4.5 La procedura di Stone come approssimazione del metodo RAS

Secondo Lavanda et al. (1996), «*usualmente, si ritiene che la precedente formula [cioè la (2.1), la formula del RAS], abbia scarso significato economico e valore statistico nullo*». Hanno quindi proposto un'interpretazione della stessa che è stata successivamente ripresa da Di Fonzo (2008).

Partendo da una nota relazione approssimata possiamo infatti interpretare la (2.1) come la correzione relativa apportata dalla stima finale $y_{i,j}$ alla stima iniziale $p_{i,j}$.

$$\log\left(\frac{y_{i,j}}{p_{i,j}}\right) \cong \frac{y_{i,j} - p_{i,j}}{p_{i,j}}$$

da cui

$$y_{i,j} \cong p_{i,j} \left(1 + \log\frac{y_{i,j}}{p_{i,j}}\right) \Rightarrow (y_{i,j} - p_{i,j}) \cong p_{i,j} \log\frac{y_{i,j}}{p_{i,j}}.$$

Sostituendo l'espressione sopra nella funzione obiettivo del RAS ne segue che

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^m y_{i,j} \log\frac{y_{i,j}}{p_{i,j}} &\cong \sum_{i=1}^n \sum_{j=1}^m p_{i,j} \left(1 + \log\frac{y_{i,j}}{p_{i,j}}\right) \log\left(\frac{y_{i,j}}{p_{i,j}}\right) = \\ &= \sum_{i=1}^n \sum_{j=1}^m \left[p_{i,j} \log\left(\frac{y_{i,j}}{p_{i,j}}\right) + p_{i,j} \left(\log\frac{y_{i,j}}{p_{i,j}}\right)^2 \right] \cong \\ &\cong \sum_{i=1}^n \sum_{j=1}^m y_{i,j} - \sum_{i=1}^n \sum_{j=1}^m p_{i,j} + \sum_{i=1}^n \sum_{j=1}^m \left[p_{i,j} \left(\log\frac{y_{i,j}}{p_{i,j}}\right)^2 \right] \end{aligned}$$

Si osservi a questo punto la terza riga: sia la prima che la seconda sommatoria sono costanti, perché corrispondono, rispettivamente, alla somma dei vincoli ed alla somma dei dati preliminari. Ne segue che minimizzare la (2.1) equivale, approssimativamente, a minimizzare

$$\sum_{i=1}^n \sum_{j=1}^m \left[p_{i,j} \left(\log\frac{y_{i,j}}{p_{i,j}}\right)^2 \right] \cong \sum_{i=1}^n \sum_{j=1}^m \left[p_{i,j} \left(\frac{y_{i,j} - p_{i,j}}{p_{i,j}}\right)^2 \right] = \sum_{i=1}^n \sum_{j=1}^m \frac{(y_{i,j} - p_{i,j})^2}{p_{i,j}}$$

dove l'ultima espressione è la funzione obiettivo della procedura di Deming e Stephan (1940) e Friedlander (1961).

2.5 La procedura DSS

Nel primo paragrafo di questo capitolo si sono visti due problemi che si possono presentare nel bilanciamento di matrici; finora si è però analizzato solo il primo. Per completezza si fornisce ora una possibile soluzione anche alla seconda situazione, quando cioè si vogliano bilanciare i totali di riga e di colonna. A differenza dei precedenti, questo caso non è propedeutico per il problema centrale di questa tesi e viene quindi presentato molto brevemente. Per approfondimenti, si veda, ad esempio, Schneider (1989a, b).

Il metodo qui proposto è il *Diagonal Similarity Scaling* (DSS) ed è analogo alla procedura RAS per il problema 1. Intuitivamente si può affermare che l'algoritmo individua la coppia riga-colonna che più delle altre viola la condizione che i totali di riga e di colonna siano uguali. La riga e la colonna individuate sono, rispettivamente, moltiplicate e divise per una costante che ne rende uguali i totali, e la procedura ricomincia.

L'algoritmo viene presentato, in termini più formali, nell'appendice B.2.

Si dimostra che, per una qualsiasi matrice non negativa P , la sequenza $\{Y^k\}$ generata dall'algoritmo converge all'unica matrice bilanciata Y che minimizza

$$f(y, p) = \sum_{j=1}^n \sum_{i=1}^n y_{ij} \left[\ln \left(\frac{y_{ij}}{p_{ij}} \right) - 1 \right]$$

rispettando il vincolo

$$\sum_{j=1}^n y_{ij} = \sum_{i=1}^n y_{ij} \quad \text{per } i, j = 1, \dots, n.$$

3 Il benchmarking di serie storiche

In questo capitolo si introduce il problema presentando il caso più semplice, di benchmarking di una serie storica. Il primo paragrafo presenta l'approccio basato sulla minimizzazione quadratica, proposto da Denton nel 1971. Il suo inserimento in questa tesi ha uno scopo didattico di introduzione al problema; essendo la procedura ormai consolidata nella prassi. Nel secondo paragrafo verranno le varianti classiche della procedura di Denton; in quello successivo si considerano le versioni proporzionali delle varie tecniche. Nel quarto si considera il principio di conservazione del movimento con riferimento al saggio di variazione. L'ultimo paragrafo propone quindi un breve riassunto delle tecniche basate sul metodo proposto da Denton.

3.1 L'approccio basato sulla minimizzazione quadratica

L'idea, già presentata, è quella che i dati sub-annuali siano uguali, in media o in somma, al totale annuale proveniente da un'altra serie, considerata nota. Una semplice soluzione può essere quella di ripartire la differenza in maniera uguale o proporzionale tra i diversi periodi (mesi o trimestri); i problemi nascono se le discrepanze non sono uniformi nei diversi anni: si crea infatti una discontinuità che non rispetta la dinamica della serie tra l'ultimo periodo di un anno e il primo di quello successivo.

La procedura descritta propone una soluzione basata su una minimizzazione quadratica vincolata, e si deve a Denton (1971). La serie storica originaria copre N anni ed è dunque formata da $n = Ns$ osservazioni, con s numero di periodi intrannuali (per esempio, se i dati sono trimestrali $s = 4$). I valori da aggiustare sono rappresentati tramite un vettore colonna p ; i totali annuali disponibili, da un'altra fonte, sono rappresentati dal vettore colonna z_t :

$$p = [p_1 \quad p_2 \quad \cdots \quad p_n]'$$

$$z_t = [z_{t_1} \quad z_{t_2} \quad \cdots \quad z_{t_N}]'$$

Il problema consiste nell'aggiustare questi valori per ottenere un vettore y con un metodo che (a) minimizzi lo scostamento dalla serie originaria e (b) soddisfi la condizione che le somme annuali dei valori della nuova serie coincidano con i valori della serie annuale. In maniera più formale, si può specificare una funzione criterio (o funzione di perdita, *penalty function*) $f(y, p)$ ed esprimere il problema in termini di scelta della y che minimizza $f(y, p)$ con il vincolo $Jy = z_t$, dove J è una matrice di aggregazione temporale, $(N \times n)$,

$$J = \begin{bmatrix} \underline{1} & \underline{0} & \cdots & \underline{0} \\ \underline{0} & \underline{1} & \cdots & \underline{0} \\ \vdots & \vdots & \ddots & \vdots \\ \underline{0} & \underline{0} & \cdots & \underline{1} \end{bmatrix}$$

con $\underline{1}$ vettore riga s -dimensionale con tutti gli elementi uguali a 1 e $\underline{0}$ vettore riga s -dimensionale nullo. Si consideri la classe delle funzioni di perdita rappresentate da $(y - p)'M(y - p)$, forma quadratica degli scarti tra i valori della serie temporale originale e quella corretta, con M matrice $(n \times n)$ non singolare definita in seguito.

Come già visto per la procedura di Stone, la soluzione è la minimizzazione della funzione lagrangiana

$$L = (y - p)'M(y - p) + 2\lambda'(Jy - z_t).$$

Il primo addendo può essere scritto sviluppando il quadrato:

$$(y - p)'M(y - p) = y'My - 2p'My + p'Mp$$

e la ricerca del minimo passa attraverso la soluzione del sistema

$$\begin{cases} \frac{\partial L}{\partial y} = 0 \\ \frac{\partial L}{\partial \lambda} = 0 \end{cases} \Rightarrow \begin{cases} 2My - 2Mp + 2J\lambda = 0 \\ Jy - z_t = 0 \end{cases} \Rightarrow \begin{bmatrix} M & J' \\ J & 0 \end{bmatrix} \begin{bmatrix} y \\ \lambda \end{bmatrix} = \begin{bmatrix} Mp \\ z_t \end{bmatrix}.$$

A questo punto è evidente che la soluzione generale, sempre valida, è

$$\begin{bmatrix} y \\ \lambda \end{bmatrix} = \begin{bmatrix} M & J' \\ J & 0 \end{bmatrix}^{-1} \begin{bmatrix} Mp \\ z_t \end{bmatrix} \quad (3.1)$$

Se M è di pieno rango, come abbiamo assunto, sfruttando un noto lemma di inversione di una matrice a blocchi, la soluzione rispetto a y si può ottenere facilmente:

$$\begin{bmatrix} M & J' \\ J & 0 \end{bmatrix}^{-1} = \begin{bmatrix} M^* & J^{*'} \\ J^* & 0 \end{bmatrix}$$

con

$$M^* = M^{-1} + M^{-1}J'(0 - JM^{-1}J')^{-1}JM^{-1}$$

$$J^{*'} = -M^{-1}J'(0 - JM^{-1}J')^{-1}$$

da cui la soluzione

$$\begin{aligned} y &= M^*Mp + J^{*'}z_t = \\ &= [M^{-1} + M^{-1}J'(0 - JM^{-1}J')^{-1}JM^{-1}]Mp + [-M^{-1}J'(0 - JM^{-1}J')^{-1}]z_t = \\ &= M^{-1}Mp - M^{-1}J'(JM^{-1}J')^{-1}JM^{-1}Mp + M^{-1}J'(JM^{-1}J')^{-1}z_t = \\ &= p + [M^{-1}J'(JM^{-1}J')^{-1}](z_t - Jp). \end{aligned}$$

La soluzione che rende minima la funzione di criterio è, come in (2.4),

$$\hat{y} = p + M^{-1}J'(JM^{-1}J')^{-1}(z_t - Jp) \quad (3.2)$$

I valori aggiustati sono quindi uguali alla somma tra i valori preliminari e una combinazione lineare delle differenze tra i totali annuali delle due serie.

3.2 Scelta della matrice M

Si inizi ora a considerare la forma della matrice M e ipotizziamo, inizialmente, che M sia la matrice identità. Ciò significa che si minimizza la somma dei quadrati delle differenze tra la serie originale e i valori aggiustati. Questo caso, che possiamo chiamare *benchmarking semplice*, implica che la funzione di perdita è minimizzata nel distribuire le discrepanze di ogni anno in misura uguale tra i diversi periodi. La sua funzione criterio, espressa in forma estesa, è

$$f(y, p) = \sum_{t=1}^n (y_t - p_t)^2. \quad (3.3)$$

Questa funzione di perdita si limita a considerare ciascun anno isolato dagli altri e genera quindi degli scalini spesso ingiustificati tra l'ultimo periodo di un anno e il primo di quello successivo: per questo $M = I$ non è una buona specificazione.

Pur soffrendo dello stesso problema del benchmarking semplice, il *benchmarking pro-rata* trova un utilizzo diffuso nella pratica corrente, perché consente in maniera semplice di distribuire la discrepanza in modo da tenere conto dell'ordine di grandezza di ciascun valore. La matrice M assume la forma $M = P^{-1}$, dove P è la matrice i cui elementi sulla diagonale sono le componenti del vettore p . La funzione di perdita è

$$f(y, p) = \sum_{t=1}^n \frac{(y_t - p_t)^2}{p_t}. \quad (3.4)$$

Si prenda ora in considerazione l'utilizzo di una *penalty function* basata sulle differenze tra le differenze prime all'indietro (*backward differences*) della serie originale ed aggiustata¹:

$$f(y, p) = \sum_{t=2}^n (\Delta y_t - \Delta p_t)^2 = \sum_{t=2}^n [\Delta(y_t - p_t)]^2. \quad (3.5)$$

La matrice M assume la forma $M = D_*' D_*$, con

$$D_* = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}. \quad (3.6)$$

di dimensioni $[(n - 1) \times n]$.

La matrice M è quindi singolare e non si possono più esprimere le stime bilanciate secondo la (3.2). Si usa allora un'approssimazione della funzione di perdita, così come presentato già da Denton (1971):

¹ Secondo la notazione di Sjöberg (1982), che verrà utilizzata nel seguito, questo metodo è indicato con la sigla D1, mentre la sua versione proporzionale è indicata con la sigla D4.

$$f(y, p) = \sum_{t=1}^n (\Delta y_t - \Delta p_t)^2 = \sum_{t=1}^n [\Delta(y_t - p_t)]^2 \quad (3.7)$$

dove Δ è l'operatore differenza ($\Delta y_t = y_t - y_{t-1}$). La funzione di perdita include il termine $\Delta(y_1 - p_1) = (y_1 - p_1) - (y_0 - p_0)$, dove il pedice 0 si riferisce all'ultimo periodo dell'anno precedente il primo anno considerato. y_0 e p_0 sono all'esterno del range nel quale la serie deve essere aggiustata: li possiamo considerare uguali e pertanto non effettuare alcun aggiustamento nella serie per gli anni all'esterno del range², da 1 (incluso) a n . Il vettore delle differenze prime (all'indietro) è indicato con $D(y - p)$, dove D è una matrice ($n \times n$)

$$D = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

e la matrice M è $M = D'D$.

La funzione di perdita può anche essere espressa in termini di differenze di ordini superiori: la funzione assume la forma più generale

$$f(y, p) = \sum_{t=1}^n (\Delta^h y_t - \Delta^h p_t)^2 = \sum_{t=1}^n [\Delta^h(y_t - p_t)]^2$$

dove h è l'operatore differenza h -esima. Eliminando dalla funzione di perdita tutti i valori esterni dal range considerato (supponendo cioè $y_t = p_t$ per ogni $t \leq 0$) il vettore delle differenze h -esime è dato da h applicazioni successive della matrice D e la matrice M assume la forma

$$M = \underbrace{D'D' \dots D'}_h \underbrace{D \dots DD}_h.$$

Nella pratica si considera la versione differenze prime, nota con l'acronimo inglese AFD (*Additive First Differences*) e la versione differenze seconde (ASD, *Additive Second Differences*)

² Questo rende vera l'ipotesi di non singolarità della matrice M e le soluzioni (approssimate) riportate. Usando un'altra notazione, si può scrivere $(y_1 - p_1)^2 + \sum_{t=2}^n [\Delta(y_t - p_t)]^2$.

Uno dei vantaggi dell'approccio sin qui illustrato è dato dalla convenienza di lavorare con questo tipo di matrici. Sia R una matrice triangolare superiore nella quale ogni elemento sulla e sopra la diagonale è 1, e gli elementi sotto la diagonale sono 0. Si dimostra facilmente che $D^{-1} = R'$ e che $(D'D)^{-1} = R'R$, $(D'D'DD)^{-1} = R'(D'D)^{-1}R$, ecc. Questo permette di non dover invertire la matrice M , che essendo $(n \times n)$ può essere molto grande, e di ottenere l'inversa di A in maniera molto semplice. Per esempio, se $h = 1$, $A = D'D$ e l'inversa è immediata:

$$M^{-1} = (D'D)^{-1} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 2 & \cdots & 2 \\ 1 & 2 & 3 & \cdots & 3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 3 & \cdots & n \end{bmatrix}.$$

Naturalmente occorre poi calcolare l'inversa della matrice $JM^{-1}J'$, che ha dimensioni $(N \times N)$. Questo calcolo richiede un impegno meno elevato: per esempio, nel caso si stesse lavorando con una serie venticinquennale di dati mensili la matrice M sarebbe 300×300 , mentre la matrice $JM^{-1}J'$ sarebbe solo 25×25 .

Un altro importante vantaggio del lavorare con le differenze delle discrepanze tra serie originale e serie preliminare è quello di ovviare al problema visto nel benchmarking semplice, che crea degli scalini ingiustificati tra l'ultimo periodo di un anno e il primo del successivo. Per questo la tecnica è nota come *principio di conservazione del movimento*.

3.3 La variante proporzionale

La funzione di perdita può anche essere espressa in termini di differenze proporzionali tra la serie originale e la serie aggiustata, anziché come differenze aritmetiche. Rispetto a quanto proposto sinora le differenze per l'applicazione sono minime: è sufficiente considerare, anziché la matrice M , la nuova matrice $Q = P^{-1}MP^{-1}$. La soluzione generale è:

$$\hat{y} = p + Q^{-1}J'(JQ^{-1}J')^{-1}(z_t - Jp) = p + PM^{-1}PJ'(JPM^{-1}PJ')^{-1}(z_t - Jp).$$

La versione proporzionale del benchmarking semplice è basata sulla matrice $Q = P^{-1}MP^{-1} = P^{-1}P^{-1}$, cioè sulla funzione criterio

$$f(y, p) = \sum_{t=1}^n \left(\frac{y_t - p_t}{p_t} \right)^2, \quad (3.8)$$

mentre la variante proporzionale della versione differenze di ordine h si propone di minimizzare la funzione

$$f(y, p) = \sum_{t=1}^n \left[\Delta^h \left(\frac{y_t - p_t}{p_t} \right) \right]^2,$$

cioè nel caso $h = 1$, noto come PFD (*Proportional First Differences*)

$$f(y, p) = \sum_{t=1}^n \left[\Delta \left(\frac{y_t - p_t}{p_t} \right) \right]^2 = \sum_{t=1}^n \left(\frac{y_t}{p_t} - \frac{y_{t-1}}{p_{t-1}} \right)^2, \quad (3.9)$$

mentre nel caso $h = 2$ (PSD, *Proportional Second Differences*)

$$f(y, p) = \sum_{t=1}^n \left[\Delta^2 \left(\frac{y_t - p_t}{p_t} \right) \right]^2 = \sum_{t=1}^n \left[\left(\frac{y_t}{p_t} - \frac{y_{t-1}}{p_{t-1}} \right) - \left(\frac{y_{t-1}}{p_{t-1}} - \frac{y_{t-2}}{p_{t-2}} \right) \right]^2.$$

Nel suo lavoro, Denton (1971) propone un esempio, aggiustando una serie quinquennale di dati trimestrali (inventati). Lo stesso esempio è riportato, insieme con lo script usato per risolverlo, nell'appendice C.2.

3.4 Riferimento al saggio di variazione nel principio di conservazione del movimento

Finora si è fatto riferimento al principio di conservazione del movimento di Denton che mira ad ottenere serie aggiustate che modificano il meno possibile i livelli della serie preliminare. In molte situazioni concrete può però risultare appropriato riferirsi ad un principio di conservazione del movimento espresso in termini di saggi di variazione (o tassi di crescita) per ottenere delle stime aggiustate che alterino il meno possibile i saggi di variazione.

Il primo criterio a cui fare riferimento è la versione basata sui tassi di crescita del criterio del *benchmarking semplice proporzionale*, ossia

$$\sum_{t=1}^n \left[\left(\frac{y_t - y_{t-1}}{y_{t-1}} \right) - \left(\frac{p_t - p_{t-1}}{p_{t-1}} \right) \right]^2 = \sum_{t=1}^n \left(\frac{y_t}{y_{t-1}} - \frac{p_t}{p_{t-1}} \right)^2.$$

Bloem et al. (2001) indicano questo criterio, seguendo Sjöberg (1982), con la sigla D3 e lo attribuiscono a Helfand et al. (1977). Questi ultimi hanno indicato anche una versione logaritmica del criterio, nota con la sigla D2, ossia:

$$\sum_{t=1}^n \left(\ln \frac{y_t}{y_{t-1}} - \ln \frac{p_t}{p_{t-1}} \right)^2 = \sum_{t=1}^n [(\ln y_t - \ln p_t) - (\ln y_{t-1} - \ln p_{t-1})]^2.$$

Appare evidente come il criterio D2 sia un'approssimazione del principio di conservazione del movimento espresso dal criterio D3.

Si consideri ora la differenza relativa tra i saggi di variazione della serie aggiustata e preliminare,

$$\sum_{t=1}^n \left[\frac{\frac{y_t - y_{t-1}}{y_{t-1}} - \frac{p_t - p_{t-1}}{p_{t-1}}}{\frac{p_t - p_{t-1}}{p_{t-1}}} \right]^2 = \sum_{t=1}^n \left(\frac{\frac{y_t}{y_{t-1}} - \frac{p_t}{p_{t-1}}}{\frac{p_t}{p_{t-1}} - 1} \right)^2,$$

oppure la sua versione approssimata

$$\sum_{t=1}^n \left[\frac{\ln \frac{y_t}{y_{t-1}} - \ln \frac{p_t}{p_{t-1}}}{\ln \frac{p_t}{p_{t-1}}} \right]^2 = \sum_{t=1}^n \left(\frac{\ln \frac{y_t}{y_{t-1}}}{\ln \frac{p_t}{p_{t-1}}} - 1 \right)^2.$$

Le quantità a denominatore dei criteri possono assumere valori nulli o molto vicini allo zero, nel qual caso il criterio rischia di “esplodere” o dar luogo a soluzioni poco accurate.

Come misure delle discrepanze tra la dinamica delle due serie si utilizza allora lo scarto da 1 dei rapporti $\left(\frac{y_t}{y_{t-1}} \right) / \left(\frac{p_t}{p_{t-1}} \right)$ e si minimizza la somma dei quadrati di tali quantità:

$$\sum_{t=1}^n \left[\frac{y_t/y_{t-1}}{p_t/p_{t-1}} - 1 \right]^2 = \sum_{t=1}^n \left[\frac{y_t/p_t}{y_{t-1}/p_{t-1}} - 1 \right]^2.$$

Questo criterio, che, seguendo ancora una volta Sjöberg (1982), è indicato con D5, è stato proposto da Skjæveland (1985)³. Usando la stessa logica già vista, si può approssimare il criterio D5 con il seguente criterio D5*:

$$\sum_{t=1}^n \left[\frac{(\ln y_t - \ln p_t) - (\ln y_{t-1} - \ln p_{t-1})}{\ln p_t - \ln p_{t-1} + 1} \right]^2.$$

Tra le varianti del metodo di Denton proposte in questo paragrafo, D5* è l'unica che permette soluzioni in forma esplicita e quindi di immediato utilizzo.

3.5 Conclusioni sul principio di conservazione del movimento di Denton

Si sono considerate sinora molte varianti del metodo di conservazione del movimento di Denton. Alcune sono state proposte da Denton stesso, altre sono state proposte successivamente e raccolte da Bloem et al. (2001) e da Di Fonzo (2003).

Riassumiamo brevemente le metodologie proposte:

- Il *benchmarking semplice* distribuisce le discrepanze tra la serie originale e la serie aggiustata; il *benchmarking pro-rata* distribuisce le discrepanze in maniera proporzionale
- L'*Additive First Differences (AFD) benchmarking* (D1) minimizza una funzione quadratica delle differenze nella crescita assoluta tra la serie originale e la serie aggiustata, cioè le differenze delle due serie differenziate di un periodo. Per estensione, Denton propone la versione generale, con le serie differenziate di h periodi, mentre Di Fonzo si limita a considerare il caso $h = 2$ (*Additive Second Differences -o ASD- benchmarking*).
- Il *benchmarking semplice proporzionale*, minimizza le differenze nella crescita proporzionale tra la serie originale e la serie aggiustata.

³ Altri autori lo attribuiscono invece a Laniel (1986).

- Il *Proportional First Differences (PFD) benchmarking* (D4) è la versione proporzionale di D1 e minimizza le differenze assolute negli aggiustamenti relativi di due periodi consecuzionali. Nel caso di serie differenziate di 2 periodi si parla invece di *Proportional Second Differences (PSD) benchmarking*.
6. La formula D2 minimizza il logaritmo delle differenze relative nei tassi di crescita delle due serie; mentre D3 minimizza le differenze assolute nei saggi di variazione. La formula D4 minimizza le differenze assolute negli aggiustamenti relativi di due periodi consecuzionali. La formula D5 minimizza le differenze relative nei tassi di crescita dei due periodi e D5* è la sua approssimazione tramite logaritmi.

Mentre tutte queste formule possono essere usate per il benchmarking, solo quelle proposte direttamente da Denton, il benchmarking pro-rata e la variante D5* hanno una formulazione esplicita dello stimatore, che è calcolabile immediatamente, senza ricorrere a procedure iterative.

4 Riconciliazione di un sistema di serie storiche

Dopo aver presentato separatamente il problema del bilanciamento di una tavola e quello del benchmarking di serie storiche univariate, si vuole adesso studiare la riconciliazione di un sistema di serie storiche. Il primo paragrafo introduce in termini formali il problema, nel caso più semplice in cui vi sia un unico vincolo contemporaneo e le stime preliminari siano collegate ad esso per somma. Nella realtà i vincoli possano essere più di uno, espressione di combinazioni lineari tra le variabili, e questo viene illustrato nel secondo paragrafo. Il terzo e il quarto paragrafo mostrano una possibile soluzione al problema: l'estensione multivariata della procedura di Denton, dovuta a Di Fonzo; mentre il quinto paragrafo ne costituisce un approfondimento, introducendo una procedura a due passi. Il sesto paragrafo presenta invece una procedura basata sul modello di regressione lineare, sviluppata da Quenneville e Rancourt (2005).

4.1 Il problema della riconciliazione

Si consideri ora il caso in cui si vogliano stimare m vettori ($n \times 1$) di dati ad alta frequenza che si riferiscono a m serie storiche di base che devono soddisfare sia il vincolo di aggregazione contemporanea che quello di aggregazione temporale.

Inizialmente le informazioni considerate note di cui si dispone sono costituite da:

- (1) z_c , vettore ($n \times 1$) di dati aggregati contemporanei;
- (2) z_{t_j} , $j = 1, \dots, m$, vettori ($N \times 1$) di dati aggregati temporalmente.

Sarà analizzata la situazione in cui siano disponibili m serie preliminari ad alta frequenza, p_j , $j = 1, \dots, m$, dove almeno uno dei due vincoli non sia soddisfatto.

Indicando con y_j , $j = 1, \dots, m$, il vettore $(n \times 1)$ di dati riconciliati che deve essere stimato, possiamo esprimere i vincoli contabili in maniera più formale:

$$\sum_{j=1}^m y_j = z_c \quad (4.1)$$

$$Jy_j = z_{t_j} \quad j = 1, \dots, m \quad (4.2)$$

dove J è una matrice di aggregazione temporale $(N \times n)$. Nel paragrafo 3.1 si è già vista una matrice di questo tipo, adesso si considera anche il caso più generale in cui ciascun z_{t_j} sia una combinazione lineare degli y_j , con coefficienti dati dal vettore j , di dimensioni $(s \times 1)$. Sia inoltre $n \geq sN$, siano cioè disponibili dei dati preliminari relativi all' $(N + 1)$ -esimo anno. La matrice J ha la forma $J = [I_N \otimes j' \quad : \quad 0]$, dove 0 è una matrice nulla $(N \times (n - sN)m)$. Nel caso presentato introducendo l'approccio univariato di Denton, $j = \underline{1}$ e $n = sN$, quindi $J = [I_N \otimes j']$.

Si riscrivano ora i vincoli in notazione matriciale: (4.1) può essere riscritta come $(\underline{1}'_m \otimes I_n)y = z_c$, essendo $\underline{1}_m$ un vettore $(m \times 1)$ di 1 e y il vettore dato da $y = [y'_1 \quad \dots \quad y'_j \quad \dots \quad y'_m]'$. Il vincolo di aggregazione temporale (4.2) può invece essere riscritto come $(I_m \otimes J)y = z_t$, con $z_t = [z'_{t_1} \quad \dots \quad z'_{t_j} \quad \dots \quad z'_{t_m}]'$. Sia H la matrice di aggregazione $((n + NM) \times nM)$

$$H = \begin{bmatrix} \underline{1}'_m \otimes I_n \\ I_m \otimes J \end{bmatrix} \quad (4.3)$$

e sia z il vettore $((n + Nm) \times nM)$, contenente sia gli aggregati contemporanei che quelli annuali, $z = [z'_c \quad z'_t]'$.

In forma matriciale possiamo indicare tutti gli $n + mN$ vincoli nella forma

$$Hy = z. \quad (4.4)$$

La presenza contemporanea di entrambi i vincoli implica l'esistenza di una relazione tra i vincoli:

$$\sum_{j=1}^m z_{t_j, T} = \sum_{h=1}^s z_{c, s(T-1)+h} \stackrel{\text{def}}{=} z_{a, T}, \quad T = 1, \dots, N$$

cioè, in forma matriciale,

$$\underline{1}'_M \otimes I_N = J_1 z, \quad (4.5)$$

dove J_1 è la matrice di aggregazione temporale¹ con $j = \underline{1}_s$.

Dalla (4.5) si può ricavare che degli $n + mN$ vincoli, soltanto $r = n + N(m - 1)$ sono liberi: già intuitivamente si può comprendere che m vincoli sono ridondanti. Formalmente è possibile scrivere la matrice H evidenziando i vincoli di aggregazione temporale che collegano y_m a z_{t_m} :

$$H = \begin{bmatrix} H_r \\ \dots \\ H_m \end{bmatrix},$$

dove $H_r = \begin{bmatrix} \underline{1}'_{m-1} \otimes I_n & I_n \\ I_{m-1} \otimes J & 0 \end{bmatrix}$ e $H_m = [0 \quad \vdots \quad J]$ sono matrici $(r \times mn)$ e $(N \times mn)$, rispettivamente. Indicando con W la matrice $(N \times r)$ $W = [J \quad \vdots \quad -(\underline{1}'_{m-1} \otimes I_N)]$ e con R la matrice $((r + N) \times r)$ $R = [I_r \quad W]'$, si ha che $H_m = WH_r$ e $H = RH_r$.

Senza perdita di generalità, si può considerare il vettore $(r \times 1)$ w , che è semplicemente il vettore z privato delle ultime M righe. Il vincolo (4.4), espresso in funzione della matrice H che non è di rango pieno, può dunque essere espresso come

$$\begin{aligned} RH_r y = Hy = z &= \begin{bmatrix} w \\ z_{t_m} \end{bmatrix} = \begin{bmatrix} z_c \\ z_{t_1} \\ \vdots \\ z_{t_{m-1}} \\ Jz - \sum_{i=1}^{m-1} z_{t_i} \end{bmatrix} = \begin{bmatrix} I_r \\ W \end{bmatrix} \begin{bmatrix} z_c \\ z_{t_1} \\ \vdots \\ z_{t_{m-1}} \end{bmatrix} \\ &= Rw. \end{aligned} \quad (4.6)$$

Considerato che R ha rango pieno di colonna allora R' è di pieno rango di riga e il prodotto $R'R$ è di pieno rango, quindi invertibile; la (4.4) può essere riscritta nella forma

$$H_r y = w. \quad (4.7)$$

¹ Nel proseguo il pedice 1 potrà essere omissso.

4.2 Vincoli contemporanei

Finora si è considerato un unico vincolo contemporaneo, contenuto nel vettore z_c , e le stime preliminari sono collegate al vettore dei vincoli esclusivamente per somma.

Adesso si introduce un caso più generale, in cui i vincoli sono più di uno, ed esprimono una qualsiasi combinazione lineare tra le variabili. Questa complicazione non è un inutile esercizio di formalismo, ma risponde alle esigenze reali, come quelle di un complesso sistema di relazioni contabili.

Iniziamo con un esempio scolastico. Siano A, B, C e D quattro diverse variabili di cui si hanno a disposizione dei valori preliminari indicati, rispettivamente, con p_A , p_B , p_C e p_D e siano X e Y due variabili che costituiscono i vincoli di aggregazione contemporanea. È noto che dev'essere $A + B + C = X$ e $B + C + D = Y$ per ogni istante temporale $i = 1, \dots, n$. Possiamo quindi formalizzare i vincoli in notazione matriciale:

$$\begin{bmatrix} y_{A,i} + y_{B,i} + y_{C,i} \\ y_{B,i} + y_{C,i} + y_{D,i} \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \Leftrightarrow \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_{A,i} \\ y_{B,i} \\ y_{C,i} \\ y_{D,i} \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix}, \quad i = 1, \dots, n.$$

Sia G la matrice di costanti note (nell'esempio, 0 e 1 ma possono essere anche -1 o pesi diversi) che definisce le relazioni tra gli y_j . Tale matrice ha dimensione $(H \times m)$, con H numero di vincoli contemporanei. L'esempio può quindi essere generalizzato nella forma

$$G \begin{bmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,j} \\ \vdots \\ y_{i,m} \end{bmatrix} = \begin{bmatrix} z_{c,i,1} \\ \vdots \\ z_{c,i,h} \\ \vdots \\ z_{c,i,H} \end{bmatrix}, \quad i = 1, \dots, n, j = 1, \dots, m, h = 1, \dots, H,$$

ovvero, usando l'usuale notazione "a dati impilati",

$$(G \otimes I_n)y = z_c,$$

con $y = [y'_1 \cdots y'_j \cdots y'_m]'$, z_{ch} , $h = 1, \dots, H$, vettori $(n \times 1)$ di dati aggregati contemporanei e $z_c = [z'_{c1} \cdots z'_{ch} \cdots z'_{cH}]'$.

La (4.3) può essere riscritta come

$$H = \begin{bmatrix} G \otimes I_n \\ I_m \otimes J \end{bmatrix}; \quad (4.8)$$

gli altri risultati visti al paragrafo precedente restano validi anche nel caso generale.

4.3 La stima della serie

Il problema diventa ora quello di stimare la serie. Si assuma che i dati disponibili siano distribuiti senza distorsione attorno al “vero” valore della serie, secondo il modello

$$p_j = \tilde{y}_j + e_j, \quad j = 1, \dots, m \quad (4.9)$$

con e_j vettore di disturbi casuali di media zero e $E(e_i e_j) = V_{ij}$, $i, j = 1, \dots, m$, con V_{ij} matrici $(n \times n)$ note. Considerando tutte le m relazioni (4.9) si ha il modello completo $p = \tilde{y} + e$, con $E(e) = 0$ e $E(ee') = V$, come già visto nel modello (2.7). Le serie bilanciate si ottengono come soluzione del problema di minimi quadrati

$$\min(y - p)' V^{-1} (y - p) \quad \text{vincolato a } Hy = z. \quad (4.10)$$

Essendo, come già detto, i vincoli (4.10) linearmente dipendenti occorre prestare attenzione nella soluzione del problema. Sia L la funzione lagrangiana

$$L = (y - p)' V^{-1} (y - p) + 2\lambda' (Hy - z)$$

con λ vettore $((r + N) \times 1)$ di moltiplicatori di Lagrange. Come in (2.4) la soluzione è data da:

$$\begin{bmatrix} V^{-1} & H' \\ H & 0 \end{bmatrix} \begin{bmatrix} y \\ \lambda \end{bmatrix} = \begin{bmatrix} V^{-1} y \\ z \end{bmatrix}.$$

Utilizzando un noto risultato² la soluzione è data da

² Si veda Magnus e Neudecker (1988).

$$\hat{y} = p + VH'(HVVH')^-(z - Hp) \quad (4.11)$$

dove $(HVVH')^-$ indica l'inversa generalizzata di Moore-Penrose di $V_a = HVH'$.

Una soluzione, equivalente alla (4.11), che non richiede l'inversione di una matrice singolare, può essere scritta in funzione degli r vincoli linearmente indipendenti. Usando la (4.7), la matrice $V_a = HVH'$ può essere riscritta come $RH_rVH_r'R' = RV_rR'$, con $V_r = H_rVH_r'$ matrice $(r \times r)$ di rango pieno.

L'inversa generalizzata di Moore-Penrose può essere costruita come segue³:

1. selezionando una sottomatrice $((r + N) \times r)$ di V , di pieno rango di colonna, chiamata S ;
2. trovando una matrice $(r \times (r + N))$, chiamata T , tale che $V = ST$;
3. l'unica inversa generalizzata di Moore-Penrose è data da:

$$V^- = T'(TT')^{-1}(S'S)^{-1}S'.$$

Scegliendo $S = RV_w$, e dunque ponendo $T = R'$, si ha che

$$V_a^- = (HVVH')^- = R(R'R)^{-1}(V_r'R'RV_r)^{-1}V_r'R'. \quad (4.12)$$

Sostituendo la (4.12) nella (4.11) si ottiene:

$$\begin{aligned} \hat{y} &= p + VH'(HVVH')^-(z - Hp) = \\ &= p + VH'R(R'R)^{-1}V_r^{-1}(R'R)^{-1}R'(z - Hp) = \\ &= p + VH_r'V_r^{-1}[(R'R)^{-1}R'z - (R'R)^{-1}R'Hp] = \\ &= p + VH_r'V_r^{-1}[(R'R)^{-1}R'Rw - (R'R)^{-1}R'RH_r p] \end{aligned}$$

cioè, semplificando,

$$\hat{y} = p + VH_r'V_r^{-1}(w - H_r p). \quad (4.13)$$

Il bilanciamento si ottiene quindi aggiungendo ai dati originari una combinazione lineare delle discrepanze. Si osservi come la (4.13) comporta l'inversione di sole matrici non singolari (evitando quindi i problemi computazionali che le inverse generalizzate provocano) e soddisfa entrambi i vincoli:

³ Si veda Di Fonzo e Marini (2003).

$$\begin{aligned}
H\hat{y} &= Hp + HVH_r'V_r^{-1}(w - H_r p) = \\
&= Hp + RH_rVH_r'V_r^{-1}(w - V_r p) = \\
&= Hp + R(w - H_r p) = Hp + z - Hp = \\
&= z.
\end{aligned}$$

4.4 Scelta della matrice V

Una scelta naturale per la matrice V è data dal principio di conservazione del movimento di Denton (1971), già presentato all'inizio del capitolo 3.

In linea con Denton e per convenienza computazionale si considera qui solo la matrice delle differenze prime approssimata. Per una trattazione delle soluzioni esatte, che considerano cioè la matrice $((n-1) \times n)$

$$\Delta = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix},$$

anziché la matrice approssimata D , si veda Di Fonzo e Marini (2006).

Si consideri ora l'estensione al caso multivariato della (3.5) (nel caso additivo) e la sua variante proporzionale:

$$\sum_{j=1}^m \sum_{t=1}^n [\Delta(y_{j,t} - p_{j,t})]^2 \quad (\text{versione additiva - AFD})$$

e

$$\sum_{j=1}^m \sum_{t=1}^n \left(\frac{y_{j,t}}{p_{j,t}} - \frac{y_{j,t-1}}{p_{j,t-1}} \right)^2 \quad (\text{versione proporzionale - PFD}). \quad (4.14)$$

In notazione matriciale possiamo scrivere le funzioni da minimizzare nella forma $\Omega_{AFD}^{-1} = I_m \otimes D'D$ e $\Omega_{PFD}^{-1} = \hat{y}^{-1} [I_m \otimes D'D] \hat{y}^{-1}$.

Ponendo $V = \Omega^{-1}$, con $\Omega = \Omega_{AFD}$ o $\Omega = \Omega_{PFD}$ a seconda della situazione, il bilanciamento si può ottenere risolvendo il problema di minimizzazione:

$$\min (y - p)' V^{-1} (y - p) \quad \text{soggetto a } Hy = z.$$

La soluzione, che segue dalla (4.13), è:

$$y = p + VH_w V_w^{-1}(w - H_w p).$$

4.5 Un approccio a due passi

L'aggiustamento simultaneo di un sistema di serie storiche soggetto sia a vincoli temporanei che contemporanei può essere difficoltoso quando il numero di variabili, periodi e vincoli è elevato. La capacità computazionale richiesta è infatti strettamente collegata alle dimensioni delle matrici coinvolte. Un'alternativa alla soluzione simultanea può essere l'approccio a due passi.

Il primo passo di questa procedura è il bilanciamento delle serie ad alta frequenza (per esempio, trimestrali) p_j alle serie annuali z_{tj} , per $j = 1, \dots, m$, usando una procedura univariata. Una soluzione appropriata in questo caso può essere la formulazione esatta del metodo di Denton nella variante PFD. Al termine di questo procedimento, si ha un insieme di dati preliminari p^* per i quali

$$(I_m \otimes J)p^* = z_t,$$

dove $p^* = [p_1^{*'} \quad \dots \quad p_j^{*'} \quad \dots \quad p_m^{*'}]$, ma

$$(G \otimes I_n)p^* \neq z_c.$$

Il passo successivo è quindi dovuto alla necessità di riconciliare p^* con le restrizioni contabili del sistema. Inizialmente si pensi ad aggiustare le variabili in ciascun periodo, senza preoccuparsi dei vincoli temporali. Dobbiamo cioè trovare y_i tale che

$$Gy_i = z_{ci},$$

$i = 1, \dots, n$, dove z_i è il vettore $(k \times 1)$ dei vincoli relativi all' i -esimo istante temporale. È ora necessario formulare il problema in modo da evitare vincoli ridondanti: in altre parole, è necessario esprimere i vincoli di aggregazione contemporanea nella forma

$$G_r y_i = w_r, \tag{4.15}$$

$i = 1, \dots, n$, dove la matrice $(r \times m)$ G_r e il vettore r -dimensionale w_r sono la forma a scala per righe (nel seguito RREF, dall'inglese *Row-Reduced Echelon Form*) di $[G \ z_i]$. I vincoli dati dalla (4.15) non sono quindi ridondanti e la matrice G_r è di rango pieno, pari a r .

Data una matrice di distanza Ω_i , con $V_i = \Omega_i^{-1}$ matrice definita positiva, dalla (2.8) si ottiene che la stima finale è data da:

$$y_i = p_i^* + V_i G_r' (G_r V_i G_r')^{-1} (w_i - G_r p_i), \quad i = 1, \dots, n. \quad (4.16)$$

Inoltre, se, come spesso accade, z_{c_i} è un vettore nullo, allora anche il vettore w_i è nullo e la (4.16) diventa

$$y_i = p_i^* - V_i G_r' (G_r V_i G_r')^{-1} G_r p_i^* = [I_m - V_i G_r' (G_r V_i G_r')^{-1} G_r] p_i^*, \quad i = 1, \dots, n.$$

Nel caso più semplice, in cui assumiamo $z_i = 0$ e $V_i = I_r$ (formulazione di Almon della procedura di minimizzazione quadratica di Stone), si ha

$$y_i = [I_m - G' (G G')^{-1} G] p_i^*, \quad i = 1, \dots, n.$$

Se si vogliono distribuire proporzionalmente le discrepanze è sufficiente utilizzare la formulazione di Deming-Stephan e Friedlander, a condizione che tutti gli elementi p siano strettamente positivi. In entrambi i casi, come già segnalato nel paragrafo 2.4.4, non si è sicuri che i segni siano preservati: può accadere che, anche in presenza di valori preliminari positivi, l'output del processo di riconciliazione produca risultati negativi.

Si consideri ora una generalizzazione che permetta di bilanciare anche i totali annuali, e si limiti per ora l'attenzione al caso dell'aggiustamento di un singolo anno. Siano

$$y_{j,k}, \quad k = 1, \dots, N,$$

gli N vettori s -dimensionali (si ricorda che s è il numero di periodi intrannuali, per esempio $s = 4$ nel caso di dati trimestrali) di osservazioni per ciascuna delle m variabili e sia y_k la vettorizzazione

$$y_k = [y'_{1,k} \quad \cdots \quad y'_{j,k} \quad \cdots \quad y'_{m,k}]'$$

Questi vettori sono vincolati dalla relazione

$$Hy_k = z_k \Leftrightarrow \begin{bmatrix} G \otimes I_s \\ I_m \otimes j' \end{bmatrix} y_k = \begin{bmatrix} z_{c,k} \\ z_{t,k} \end{bmatrix}$$

e la stima finale con la procedura di Stone è data da

$$y_k = p_k^* + V_k H_r' (H_r V_k H_r')^{-1} (w_k - H_r p_k^*) \quad (4.17)$$

per ogni $k = 1, \dots, N$.

La procedura può essere estesa, considerando l'aggiustamento simultaneo di ν anni consecutivi, con $1 < \nu < N$. Solo per semplicità, si supponga che ν sia un divisore di N , in modo che $N^* = N/\nu$ sia un numero intero. Le osservazioni di ciascuna variabile possono quindi essere raggruppate in N^* vettori:

$$y_{j,k^*} = [y_{j,\nu(k^*-1)+1} \quad y_{j,\nu(k^*-1)+2} \quad \cdots \quad y_{j,\nu k^*}]', \quad k^* = 1, \dots, N^*, \quad j = 1, \dots, m.$$

Ciascun vettore ha lunghezza $s\nu$. La forma "impilata" (*stacked form*) è il vettore ($s\nu m \times 1$)

$$y_{k^*} = [y'_{1,k^*} \quad \cdots \quad y'_{j,k^*} \quad \cdots \quad y'_{m,k^*}]'$$

Il problema di riconciliazione è esprimibile dal vincolo

$$Hy_{k^*} = z_{k^*} \Leftrightarrow \begin{bmatrix} G \otimes I_{s\nu} \\ I_m \otimes J \end{bmatrix} y_{k^*} = \begin{bmatrix} z_{c,k^*} \\ z_{t,k^*} \end{bmatrix} \quad (4.18)$$

Riformulando la (4.17) si ottiene la stima:

$$y_{k^*} = p_{k^*}^* + V_{k^*} H_r' (H_r V_{k^*} H_r')^{-1} (w_{k^*} - H_r p_{k^*}^*),$$

$k^* = 1, \dots, N^*$.

Si noti che se $\nu = N$ allora $y_{k^*} = y$ e il sistema (4.18) diventa uguale alla riconciliazione simultanea. Il vantaggio di questa formulazione è che l'aggiustamento di pochi anni alla volta (per esempio, due o tre) riduce drasticamente il tempo computazionale richiesto per la stima.

4.6 Riconciliazione: un approccio basato sulla regressione

Abbiamo finora visto una procedura per la riconciliazione di sistemi di serie storiche. Un'interessante alternativa si deve a Quenneville e Rancourt (2005) ed è basata sul modello di regressione lineare.

Si inizi pensando a tre sole osservazioni: p_1 , p_2 e z tali che la somma delle prime due sia uguale alla terza. Un modo per aggiustare i valori di p_1 e p_2 è l'aggiustamento pro-rata: il valore aggiustato di p_1 è $y_1 = p_1 z / (p_1 + p_2)$ e quello di p_2 è $y_2 = p_2 z / (p_1 + p_2)$.

In chiave statistica si ponga $p_1 = y_1 + e_1$, $p_2 = y_2 + e_2$, $z = y_1 + y_2$, dove e_1 (rispettivamente e_2) è una variabile casuale di media nulla e varianza p_1 (rispettivamente p_2). Questa modellizzazione è ridondante, e può essere semplificata eliminando y_2 : $p_1 = y_1 + e_1$, $z - p_2 = y_1 + e_2$. Assumendo che e_1 e e_2 siano tra loro incorrelati, il migliore stimatore lineare non distorto (BLUE) è una media pesata di p_1 e $z - p_2$ i cui pesi sono inversamente proporzionali alla varianza:

$$y_1 = \left(\frac{1}{p_1} + \frac{1}{p_2} \right)^{-1} \left(\frac{p_1}{p_1} + \frac{z - p_2}{p_2} \right) = p_1 \frac{z}{p_1 + p_2}. \quad (4.19)$$

Il codice Matlab® riportato nell'appendice C.3 mostra come il modello di regressione sia semplicemente uno strumento per effettuare, in pratica, l'aggiustamento pro-rata.

Anche in questo caso, la varianza può essere modificata introducendo dei coefficienti di alterabilità, che indicano quali valori siano più credibili e quindi debbano essere meno modificati dalla riconciliazione. Questo aspetto è molto importante, perché permette di considerare la riconciliazione non semplicemente come un esercizio di algebra matriciale ma come una procedura che risponde alle reali esigenze di chi pubblica e di chi utilizza i set di dati.

Un coefficiente di alterabilità può essere associato anche al vincolo, che perde quindi la caratteristica di essere stringente. L'uso di coefficienti di alterabilità vicini a

zero per il vincolo è un artificio per avvicinarsi ad un vincolo stringente, con i vantaggi di semplificazione di una regressione lineare senza vincoli nei parametri.

Nel caso di due soli valori, si ha che:

$$y_1 = p_1 + \frac{a_1 p_1}{a_z z + a_1 p_1 + a_2 p_2} [z - (p_1 + p_2)]$$

$$y_2 = p_2 + \frac{a_2 p_2}{a_z z + a_1 p_1 + a_2 p_2} [z - (p_1 + p_2)]$$

con

$$y_1 + y_2 = z - \frac{a_z z}{a_z z + a_1 p_1 + a_2 p_2} [z - (p_1 + p_2)].$$

Se l'uso della regressione lineare nel caso di un'unica serie di dati è una complicazione matematica inutile, le potenzialità di questa tecnica emergono quando si parla di sistemi di serie storiche.

Nel caso più generale, infatti, abbiamo una tabella di dimensioni $(n \times m)$, con dei vincoli non stringenti, che dà luogo a un modello di regressione lineare costituito da $nm + n + m + 1$ equazioni, dove le prime nm sono riferite a ciascun valore interno, le successive $n + m$ sono date dai totali di colonna e di riga, e l'ultima dal totale generale.

$$p_{i,j} = y_{i,j} + e_{i,j}, \quad e_{i,j} \sim (0, a_{i,j} p_{i,j})$$

$$p_{i,\cdot} = \sum_{q=1}^m y_{i,q} + e_{i,\cdot}, \quad e_{i,\cdot} \sim (0, a_{i,\cdot} p_{i,\cdot})$$

$$p_{\cdot,j} = \sum_{k=1}^n y_{k,j} + e_{\cdot,j}, \quad e_{\cdot,j} \sim (0, a_{\cdot,j} p_{\cdot,j})$$

$$p_{\cdot,\cdot} = \sum_{k=1}^n \sum_{q=1}^m y_{k,q} + e_{\cdot,\cdot}, \quad e_{\cdot,\cdot} \sim (0, a_{\cdot,\cdot} p_{\cdot,\cdot}).$$

Se uno dei coefficienti di alterabilità è pari a 0, allora la relativa equazione deve essere sostituita da un vincolo. In pratica può risultare computazionalmente molto conveniente sostituire lo 0 con un valore positivo molto piccolo.

La procedura multivariata di Quenneville e Rancourt, in pratica, è una procedura a due passi: le serie vengono prima trattate univariatamente con la procedura di Denton per preservarne la dinamica temporale, quindi si procede alla riconciliazione con la tecnica di bilanciamento di Stone.

5 Un'applicazione: l'indagine sul commercio al dettaglio

Dopo aver affrontato il problema della riconciliazione statistica di sistemi di serie storiche dal punto di vista metodologico e tecnico, si ritiene indispensabile mostrare una possibile applicazione pratica delle procedure viste fin qui. Il capitolo che segue mostra quindi l'applicazione delle procedure di riconciliazione ai dati mensili sul commercio al dettaglio in Canada, dal 1991 al 2003. I primi tre paragrafi introducono l'indagine, il quarto presenta i dati, i successivi tre presentano la riconciliazione nel caso di un vincolo esogeno. Infine l'ultimo paragrafo si occupa di riconciliare i dati in presenza di un vincolo endogeno.

5.1 Il Canada



Figura 5.1 Carta politica del Canada

Il Canada¹ è la nazione più settentrionale del continente americano; coi suoi 9.976.140 chilometri quadrati è il secondo paese più esteso del mondo. È uno stato federale, retto da una monarchia costituzionale con sistema parlamentare e il governo federale ha l'obbligo di fornire servizi nelle due lingue ufficiali del paese, l'inglese e il francese. La capitale è Ottawa, in Ontario, ma le città principali sono Toronto (in Ontario) e Montreal (in Quebec), le uniche che superano il milione di abitanti.

È suddiviso in dieci province e tre territori, come mostrato in Tabella 5.1. Il territorio del Nunavut è stato creato nel 1999 per distaccamento dai Territori del Nord-Ovest per riconoscere l'identità di patria al popolo Inuit ed è una delle aree meno densamente abitate al mondo.

Codice		Nome della provincia/territorio ^A	Capitale	Popolazione ^B		Superficie ^B	
				abitanti	%	km ²	%
RG59	P	Columbia Britannica	Victoria	4.113.487	13,01	924.815	10,26
RG48	P	Alberta	Edmonton	3.290.350	10,41	640.045	7,10
RG47	P	Saskatchewan	Regina	968.157	3,06	588.276	6,52
RG46	P	Manitoba	Winnipeg	1.148.401	3,63	552.370	6,13
RG35	P	Ontario	Toronto	12.160.282	38,47	907.574	10,06
RG24	P	Quebec	Québec	7.546.131	23,87	1.356.367	15,04
RG13	P	Nuovo Brunswick	Fredericton	729.997	2,31	71.355	0,79
RG12	P	Nuova Scozia	Halifax	913.462	2,89	52.917	0,59
RG11	P	Isola del Principe Edoardo	Charlottetown	135.851	0,43	5.684	0,06
RG10	P	Terranova e Labrador	St. John's	505.469	1,60	370.495	4,11
RG60	T	Yukon	Whitehorse	30.372	0,10	474.711	5,26
RG61	T	Territori del Nord-Ovest	Yellowknife	41.464	0,13	1.140.835	12,65
RG62	T	Nunavut	Iqaluit	29.474	0,09	1.932.255	21,43
Totale Canada			Ottawa	31.612.897	100,00	9.017.698	100,00

^A Per comodità del lettore, quando disponibile, ci si riferirà alle località geografiche con l'esonimo italiano e non con la denominazione ufficiale.

^B Fonte: Statistics Canada (censimento della popolazione, 2006).

Tabella 5.1 Province (P) e territori (T) canadesi con rispettivo codice identificativo, capitale, popolazione e superficie in km².

¹ L'inclusione di questa breve introduzione nella tesi è un omaggio dell'autore al paese che sogna fin da bambino. Le informazioni contenute possono facilitare la comprensione di alcune considerazioni successive.

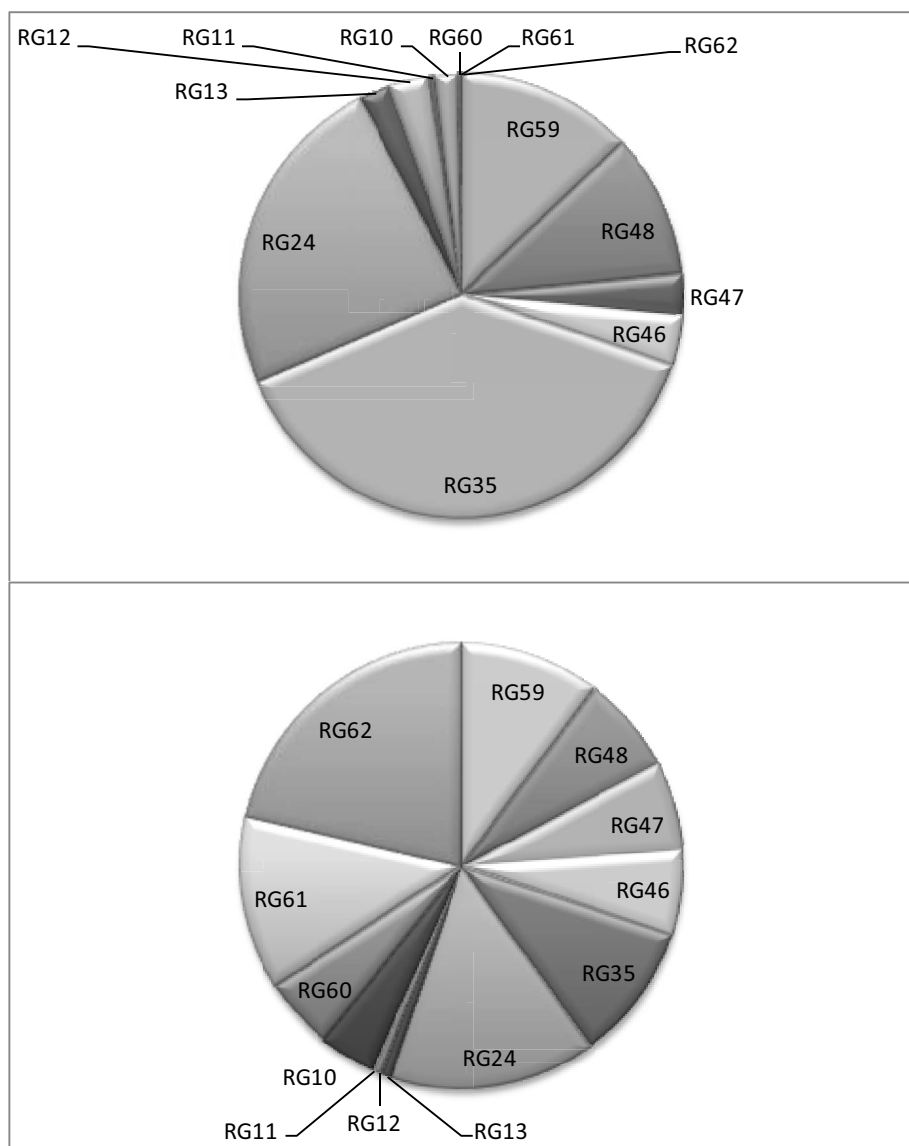


Figura 5.2 Popolazione (in alto) e superficie delle diverse province/territori.

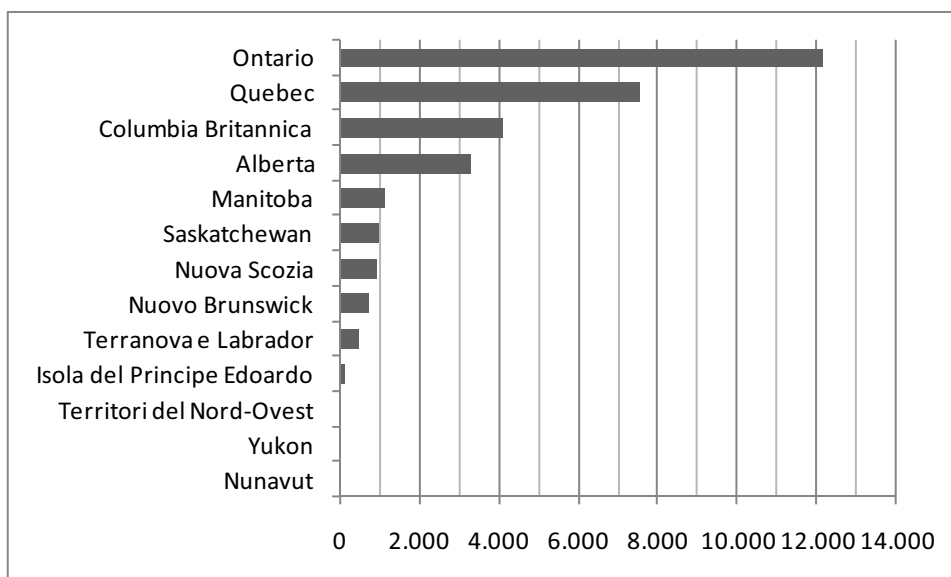


Figura 5.3 Popolazione, in migliaia di abitanti, delle province/territori.

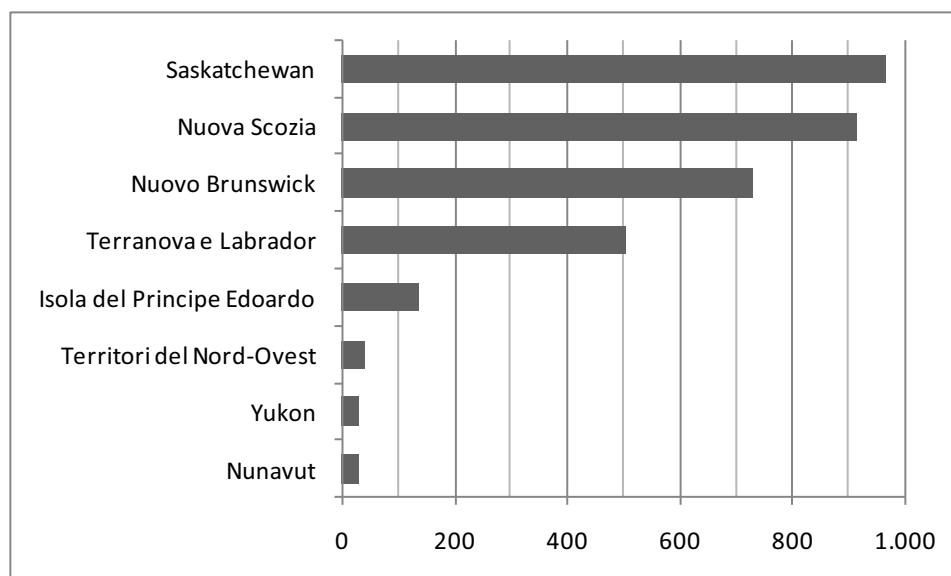


Figura 5.4 Popolazione, in migliaia di abitanti, delle province/territori con meno di un milione di abitanti.

La valuta è il Dollaro Canadese, che vale circa 0,6 euro². L'agenzia governativa federale che effettua studi statistici sul paese, anche a livello delle singole province o territori, è Statistics Canada (Statistique Canada).

5.2 L'indagine mensile sul commercio al dettaglio in Canada

“L'indagine mensile sul commercio al dettaglio (Monthly Retail Trade Survey, MRTS³) fornisce informazioni mensili sui risultati del settore delle vendite al dettaglio e, se combinato con altre statistiche, rappresenta un importante indicatore dell'economia canadese.”⁴

L'indagine MRTS fornisce una misura della salute e delle performance del settore del commercio al dettaglio canadese. Le stime di vendita ottenute dai rivenditori sono un indicatore mensile chiave dei modelli di acquisto dei consumatori in Canada. Inoltre, le vendite al dettaglio sono un'importante componente del Prodotto Interno Lordo e sono parte di molti modelli economici usati da agenzie pubbliche e private. La Bank of Canada si basa anche sulle stime mensili di vendita al dettaglio quando prende decisioni che influenzano i tassi di interesse.

² 0,6724 € al 31/01/2009. Fonte: Bank of Canada.

³ In francese Enquête sur le commerce de détail (mensuelle) (EMCD).

⁴ Tradotto da Statistics Canada (2008).

La popolazione di riferimento include tutte le unità statistiche del Registro Aziende di Statistics Canada (*Statistics Canada's Business Register*) classificate come appartenenti al settore "commercio al dettaglio" secondo la classificazione NAICS (*North American Industry Classification System*). I dati raccolti non includono nessuna forma di vendita diretta che esclude i punti vendita (vendita porta-a-porta, distributori automatici, vendite per corrispondenza, ecc.).

5.2.1 Il campionamento

Il campione considerato dall'indagine MRTS è formato da 10.000 gruppi (cluster) di unità economiche (punti vendita al dettaglio) selezionati dal Registro Imprese di Statistics Canada. Un cluster di unità economiche è dato dall'insieme dei punti vendita appartenenti ad una stessa impresa che operano nello stesso settore commerciale e nella stessa regione geografica.

L'MRTS usa un disegno di campionamento stratificato, con campionamento casuale semplice da ciascun strato. Per ciascun settore commerciale e per ciascuna regione geografica viene selezionato un certo numero di punti vendita. Nella stratificazione, le regioni geografiche sono 16: oltre alle dieci province e ai tre territori esistono tre aree metropolitane statistiche (*Census Metropolitan Area*) corrispondenti alle tre principali città canadesi (Toronto, Montreal e Ottawa) e ai loro dintorni. Successivamente i punti vendita vengono stratificati secondo la loro dimensione, ottenuta combinando dati provenienti da altre indagini e dati amministrativi (perlopiù fiscali).

Gli strati secondo la dimensione comprendono uno strato *take-all* (censimento), almeno due strati *take-some* (campionamento parziale) e uno strato *take-none* (nessun campionamento). Lo strato *take-none* serve a ridurre le risposte necessarie, escludendo una categoria di attività commerciali che, nel loro complesso, rappresentano meno del 5% delle vendite totali; anziché inviare loro i questionari questi dati vengono stimati da fonti amministrative.

5.2.2 I questionari

L'indagine MRTS è costituita da tre sotto-indagini:

- (1) l'indagine mensile sul commercio al dettaglio (*Monthly Retail Trade Survey*),
- (2) l'indagine MRTS con valutazione delle scorte (*MRTS with inventories*),
- (3) l'indagine delle vendite e degli inventari delle bevande alcoliche (*Survey of Sales and Inventories of Alcoholic Beverages*).

I questionari raccolgono mensilmente dati sulle vendite al dettaglio e sul numero di esercizi commerciali per provincia o territorio e sulle merci immagazzinate da un campione di punti vendita.

Tutti i questionari sono molto brevi (due pagine di cui una di istruzioni) e chiedono alcuni dati utili per contattare l'attività commerciale (indirizzo, nome del responsabile ecc.) e:

- (1) nel primo caso il numero di punti vendita e di vendite (e altre entrate, in dollari canadesi) in ciascuna regione geografica nel mese di riferimento;
- (2) nel secondo caso, oltre alle informazioni già richieste nel primo questionario, il valore (al costo) delle merci possedute e pronte alla vendita alla fine del mese di riferimento, indicando solo il totale e non dettagliate per ogni regione geografica;
- (3) nel terzo caso, oltre alle informazioni richieste nei primi due questionari, si vogliono sapere le vendite per tipologia di cliente (cliente finale, rivenditore ecc.), solo per le province e i territori (e non per le tre aree metropolitane). Il questionario, costruito in collaborazione con gli interessati, prevede anche uno spazio di commento.

5.2.3 Mancate risposte

Nonostante gli sforzi di Statistics Canada perché i questionari vengano compilati accuratamente, si hanno sempre delle mancate risposte. Affinché un'unità statistica che ha risposto solo parzialmente al questionario venga classificata come rispondente, deve soddisfare un livello di soglia minimo, sotto al quale l'unità statistica viene rifiutata e considerata come non rispondente.

Le non risposte hanno due effetti sui dati:

- (1) introducono distorsione nelle stime se il gruppo dei non rispondenti presenta caratteristiche diverse dai rispondenti, e
- (2) contribuiscono ad aumentare la varianza campionaria delle stime perché la numerosità effettiva del campione è minore di quella pensata originariamente.

Per ridurre l'impatto delle non risposte Statistics Canada effettua il re-invio dei questionari e telefonate alle unità del campione. Se il punto vendita non dispone ancora del dato al momento richiesto, si accetta una stima approssimativa, che sarà successivamente corretta quando sarà disponibile il dato esatto. Inoltre, per instaurare un rapporto di fiducia tra intervistatore e rispondente, a ciascuna unità del campione viene assegnato lo stesso intervistatore ogni mese.

Per ridurre invece l'effetto delle mancate risposte che, inevitabilmente, sono presenti si ricorre a due soluzioni:

- (1) aumentare il numero di unità statistiche selezionate, attraverso l'uso di tassi di *over-sampling*, e
- (2) diminuire il numero di non risposte attraverso l'imputazione dei dati mancanti.

5.2.4 Raccolta dei dati

La raccolta dei dati è effettuata dagli uffici regionali di Statistics Canada. Le unità selezionate ricevono il questionario per posta o sono contattate telefonicamente per conoscere il valore delle loro vendite e delle merci in magazzino. Le nuove unità selezionate per l'indagine ricevono una lettera di presentazione che annuncia la telefonata di un intervistatore di Statistics Canada: con questa telefonata Statistics Canada ottiene conferma del tipo di attività del rispondente e inizia a raccogliere i dati.

Per ulteriori informazioni sulle procedure di raccolta dei dati, del controllo della coerenza dei dati, dell'imputazione dei dati mancanti e sull'ottenimento delle stime, si consulti Statistics Canada (2008).

5.3 La procedura di destagionalizzazione

L'applicazione sviluppata nei prossimi paragrafi riguarda la riconciliazione statistica di dati destagionalizzati a dei benchmark temporali e contemporanei. Per questo, si ritiene opportuno ricordare brevemente le principali caratteristiche della procedura di destagionalizzazione, così come presentata da Statistics Canada (2008).

Le serie storiche socio-economiche come quelle dell'indagine MRTS possono abitualmente essere scomposte in cinque componenti principali, di seguito descritte.

- (1) Il ciclo-trend. Il trend rappresenta la tendenza di fondo del fenomeno considerato, riferita ad un lungo periodo di tempo; il ciclo rappresenta invece le fluttuazioni attribuibili a fasi di crescita e fasi di declino. Spesso queste due componenti sono stimate insieme.
- (2) La componente stagionale, costituita da fenomeni che, per influenza di fattori climatici o sociali, tendono a ripetersi in maniera pressoché analoga nel medesimo periodo di anni successivi.
- (3) Le variazioni di calendario. La diversa lunghezza dei mesi e la presenza di un diverso numero di giorni lavorativi influenzano i dati mensili.
- (4) L'effetto delle vacanze di Pasqua, che porta ad anticipare a Marzo molte attività di Aprile quando Pasqua cade in Marzo.
- (5) La componente irregolare, che include tutte le altre fluttuazioni, più o meno casuali, non comprese nelle precedenti componenti. È una componente residuale che include errori di misura nelle variabili e eventi imprevedibili (scioperi, siccità, alluvioni o altri eventi inattesi che provocano variazioni nelle attività commerciali dei rispondenti).

Le ultime quattro componenti disturbano la componente trend-ciclo della serie. La procedura di destagionalizzazione consiste nel rimuovere la componente stagionale e le altre componenti che presentano regolarità (variazioni di calendario e effetto della Pasqua). Le serie destagionalizzate, tuttavia, contengono non solo il trend-ciclo ma anche le componenti irregolari, che sono difficili da isolare.

Il metodo X11-ARIMA (Bee Dagum, 1978) è usato per destagionalizzare serie mensili e trimestrali. È basato su un principio iterativo applicato nella stima dei diversi componenti, e le stime ad ogni passo sono effettuate usando un adeguato modello ARIMA. Il metodo X-11 include un tool per rilevare e correggere punti atipici, usato per pulire le serie prima della destagionalizzazione con i modelli ARIMA.

Infine l'effetto delle variazioni di calendario e della Pasqua viene stimato usando un modello di regressione lineare con errori ARIMA, procedendo iterativamente: stima delle componenti, ricerca di effetti indesiderati sulla componente irregolare, stima delle componenti su una serie corretta, ricerca di effetti indesiderati sulla componente irregolare ecc.

5.4 La struttura dei dati analizzati

I dati utilizzati nell'applicazione⁵ sono disponibili secondo due diverse classificazioni: per provincia/territorio (PR) o per settore commerciale (TG, *trade group*). Per ciascuna classificazione sono disponibili due set di dati: i dati grezzi (indicati anche RAW, dall'equivalente inglese) e i dati destagionalizzati (SA, *seasonally adjusted*). L'insieme dei due set di dati di ciascuna classificazione, e dei benchmark descritti nel prossimo capitolo, formano un sistema.

Le serie mensili coprono 156 mesi, da Gennaio 1991 a Dicembre 2003, e sono aggregate per le tredici province/territori in un caso (Tabella 5.2) e per i diciannove settori commerciali (Tabella 5.3) nell'altro.

Le province e i territori hanno quote di mercato (Tabella 5.2 e Figura 5.5) che rispecchiano il numero di abitanti: il 60% delle vendite avviene nelle due province più abitate (Quebec e Ontario); la percentuale sale all'85% se si considerano anche Alberta e Columbia Britannica. I tre territori, nonostante si estendano per 3,5 milioni di chilometri quadrati, danno conto dello 0,5% delle vendite al dettaglio.

⁵ Il lettore interessato può richiederli contattando per e-mail l'autore.

Codice		Nome della provincia/territorio	Quota di mercato media
RG10	P	Terranova e Labrador	1,68%
RG11	P	Isola del Principe Edoardo	0,43%
RG12	P	Nuova Scozia	3,17%
RG13	P	Nuovo Brunswick	2,49%
RG24	P	Quebec	22,82%
RG35	P	Ontario	37,87%
RG46	P	Manitoba	3,33%
RG47	P	Saskatchewan	2,98%
RG48	P	Alberta	10,92%
RG59	P	Columbia Britannica	13,94%
RG60	T	Yukon	0,12%
RG61	T	Territori del Nord-Ovest	0,18%
RG62	T	Nunavut	0,09%

Tabella 5.2 Province e territori del Canada, relativi codici identificativi e quote di mercato medie.

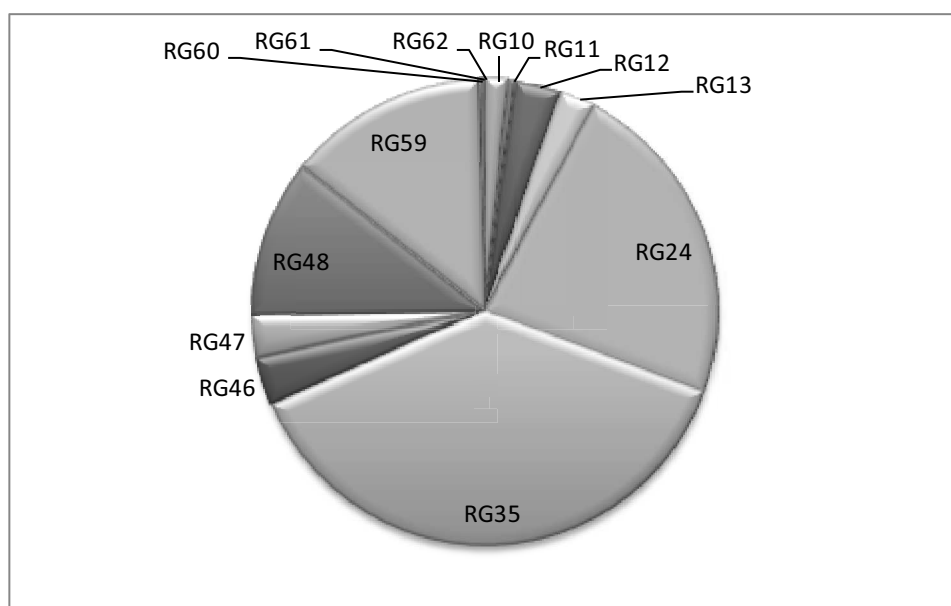


Figura 5.5 Quote di mercato medie delle diverse province/territori.

Codice	Settore commerciale	Quota di mercato media
TG01	<i>New Car Dealers</i> Venditori di auto nuove	19,98%
TG02	<i>Used and Recreational Motor Vehicle and Parts Dealers</i> Venditori di auto usate e di pezzi di ricambio	4,17%
TG03	<i>Furniture Stores</i> Negozi di mobili	2,29%
TG04	<i>Home Furnishings Stores</i> Negozi di articoli per la casa	1,08%
TG05	<i>Computer and Software Stores</i> Negozi di computer e software	0,62%
TG06	<i>Home Electronics and Appliance Stores</i> Negozi di elettrodomestici ed elettronica	2,37%
TG07	<i>Home Centres and Hardware Stores</i> Ferramenta	3,42%
TG08	<i>Specialized Building Materials and Garden Stores</i> Negozi di materiale per l'edilizia ed il giardinaggio	1,30%
TG09	<i>Supermarkets</i> Supermercati	18,74%
TG10	<i>Convenience and Specialty Food Stores</i> Gastronomie e negozi di prodotti tipici	2,82%
TG11	<i>Beer, Wine and Liquor Stores</i> Negozi di birra, vino e liquori	3,84%
TG12	<i>Pharmacies and Personal Care Stores</i> Farmacie e negozi per la cura personale	6,82%
TG13	<i>Gasoline Stations</i> Stazioni di servizio	8,62%
TG14	<i>Clothing Stores</i> Negozi di abbigliamento	4,79%
TG15	<i>Shoe, Clothing Accessories and Jewellery Stores</i> Negozi di scarpe, accessori e gioiellerie	1,73%
TG16	<i>Sporting Goods, Hobby, Music and Book Stores</i> Negozi di articoli sportivi, di hobbistica, libri e musica	2,75%
TG17	<i>Department Stores</i> Grandi magazzini	6,27%
TG18	<i>Other General Merchandise Stores</i> Altri negozi generalisti	5,83%
TG19	<i>Miscellaneous Store Retailers</i> Altri venditori al dettaglio	2,57%

Tabella 5.3 Settori commerciali considerati nell'indagine MRTS, relativi codici identificativi e quote di mercato medie.

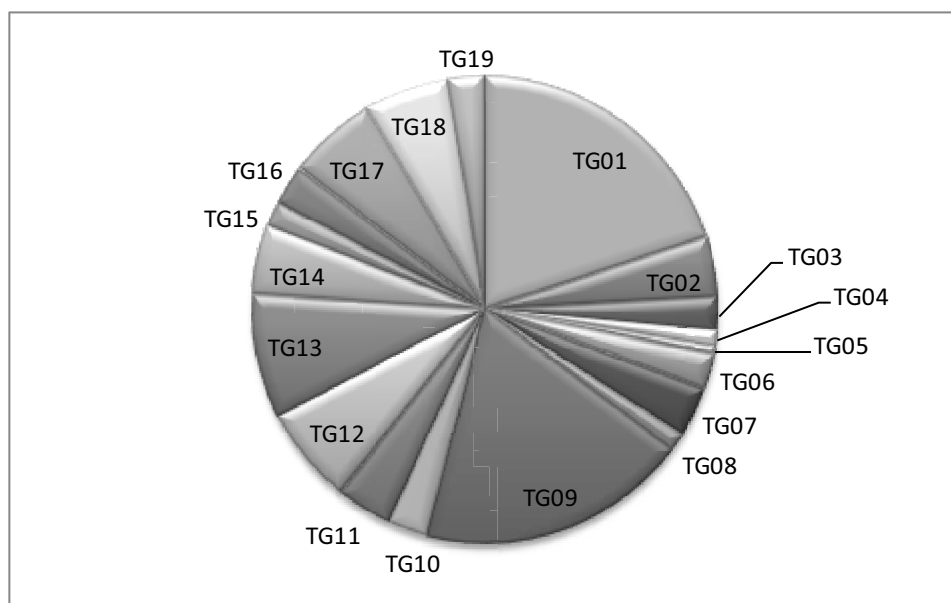


Figura 5.6 Quote di mercato medie dei diversi settori commerciali.

Come si può osservare (Tabella 5.3 e Figura 5.6), i settori commerciali hanno quote di mercato medie molto diverse tra loro. I venditori di auto nuove (TG01) e i supermercati (TG09) sono gli unici a superare il 10% e, uniti, danno conto di quasi quattro dollari ogni 10 di vendite. Una seconda fascia di settori commerciali è costituita dalle farmacie e negozi per la cura personale (TG12), dalle stazioni di servizio (TG13), dai *Department Stores* (grandi magazzini, TG17) e da *Other General Merchandise Store* (TG18), che costituiscono singolarmente tra il 5 e il 10% delle vendite. Le altre 13 categorie, complessivamente, rappresentano circa un terzo delle vendite in Canada. I grafici che seguono mostrano le serie storiche delle vendite (grezze e destagionalizzate), sia per provincia che per settore commerciale.

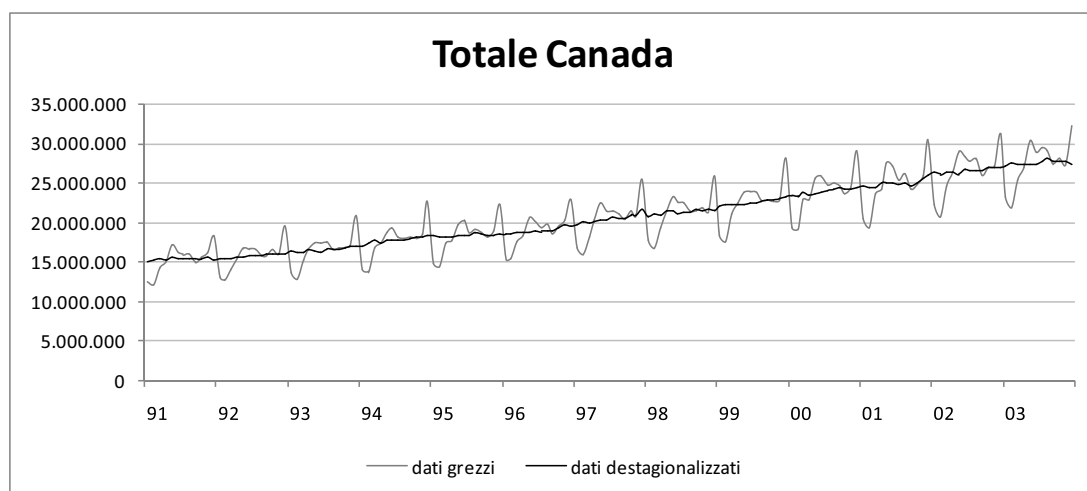
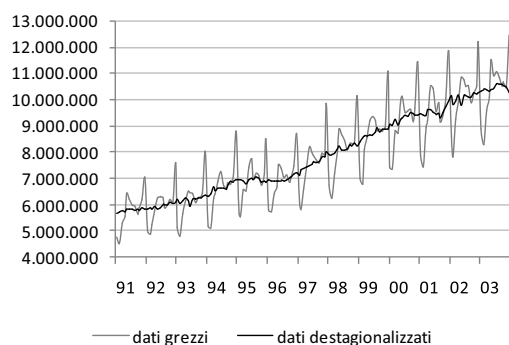
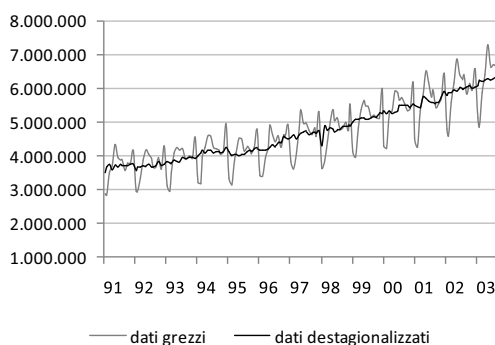
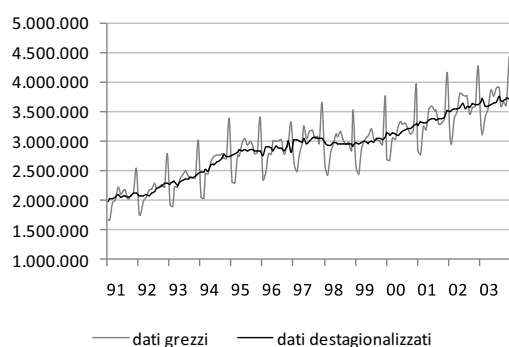
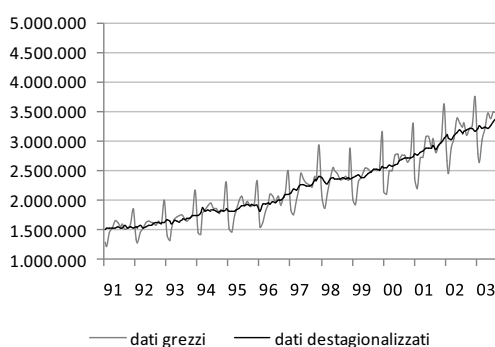
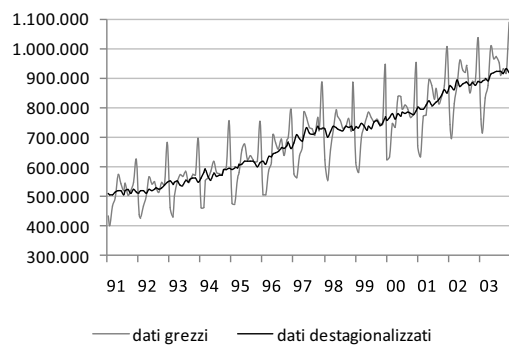
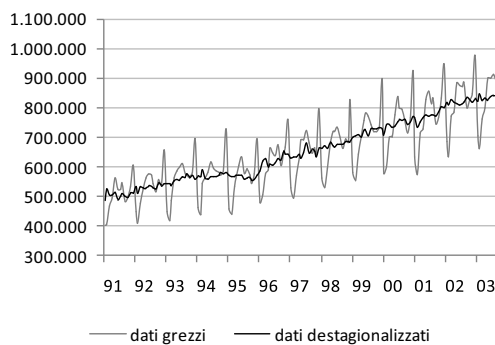
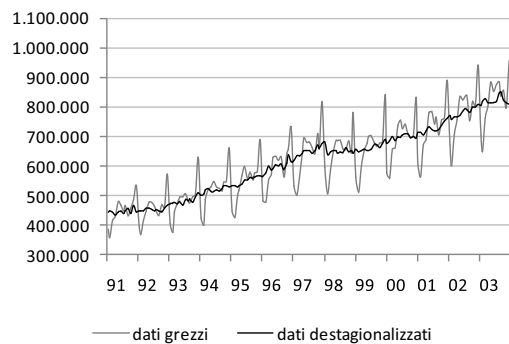
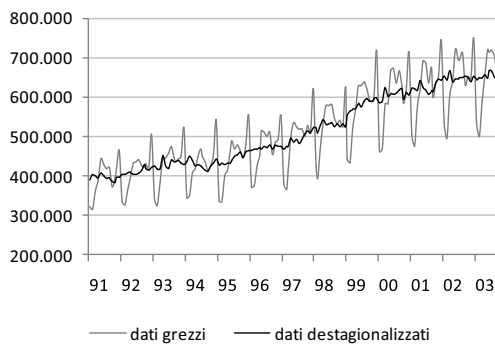


Figura 5.7 Vendite mensili al dettaglio in Canada da gennaio 1991 a dicembre 2003. Dati destagionalizzati indirettamente dalla somma delle serie destagionalizzate per settore commerciale.

Ontario**Quebec****Columbia Britannica****Alberta****Manitoba****Nuova Scozia****Saskatchewan****Nuovo Brunswick**

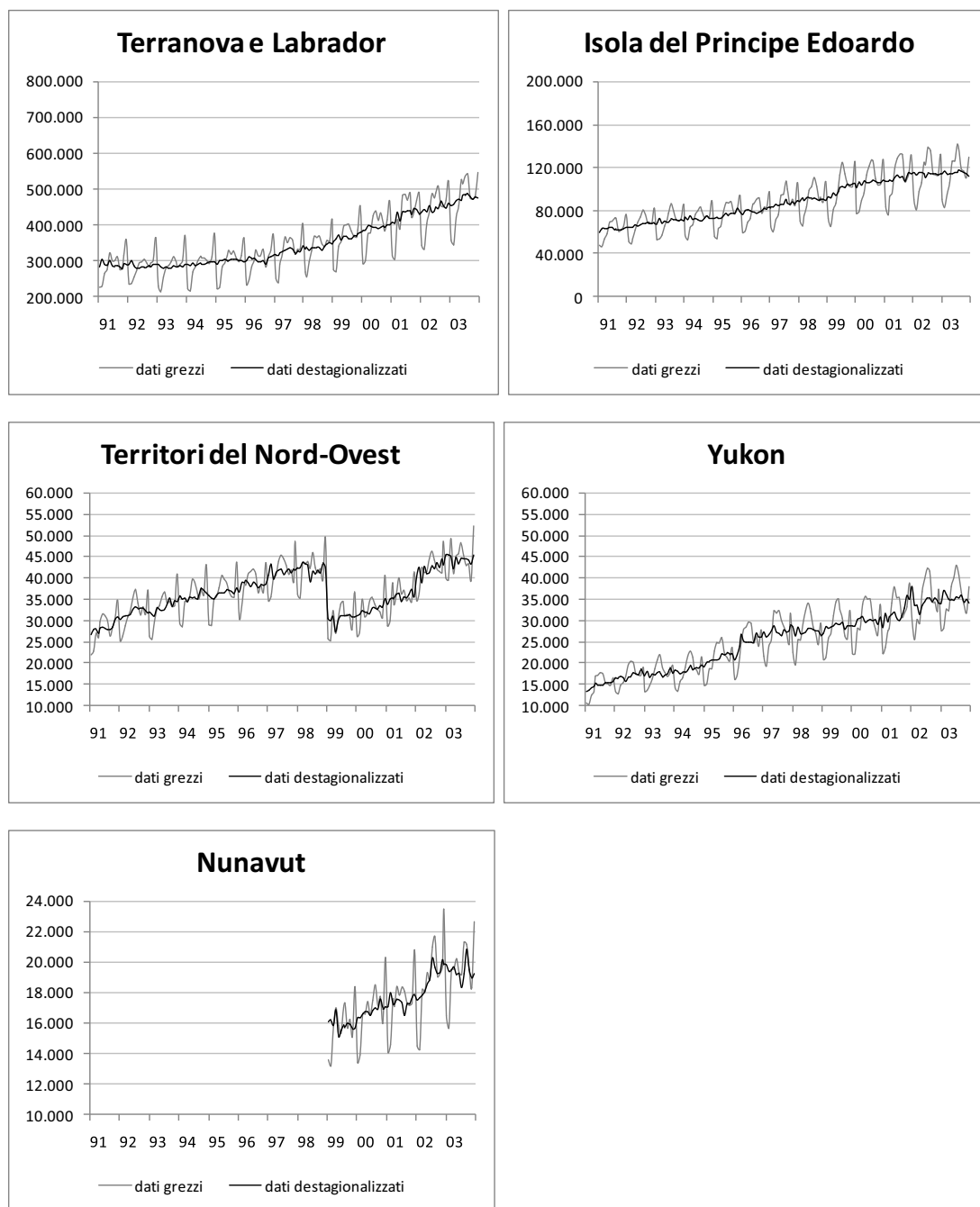
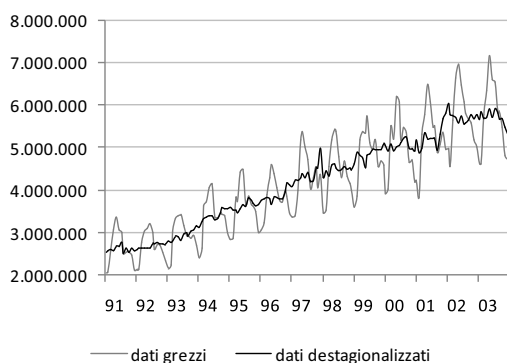
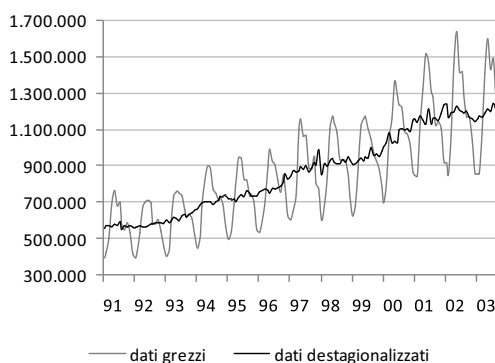
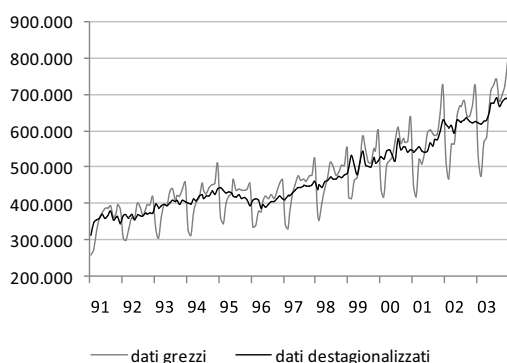
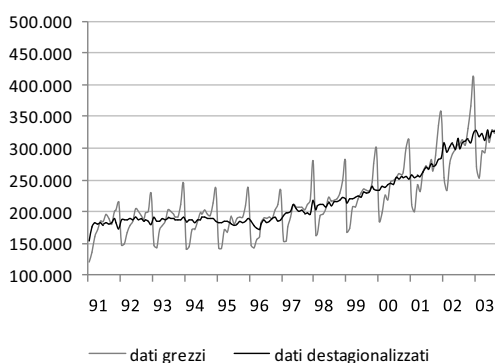
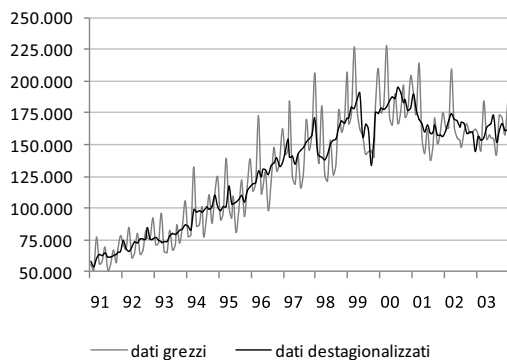
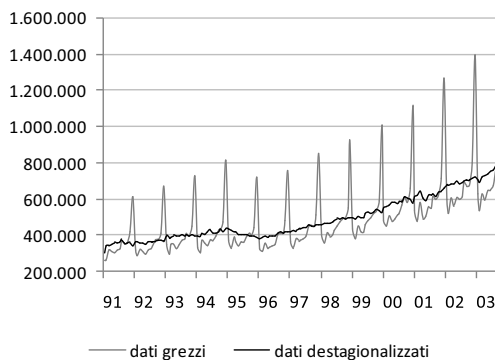
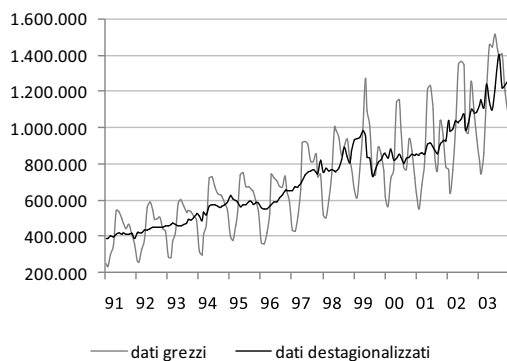
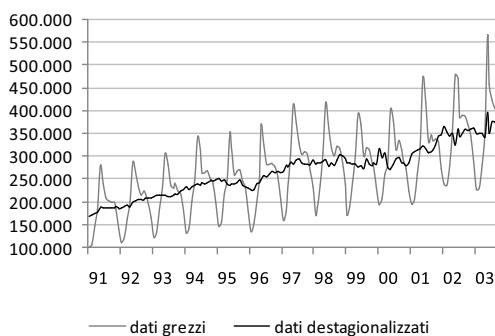
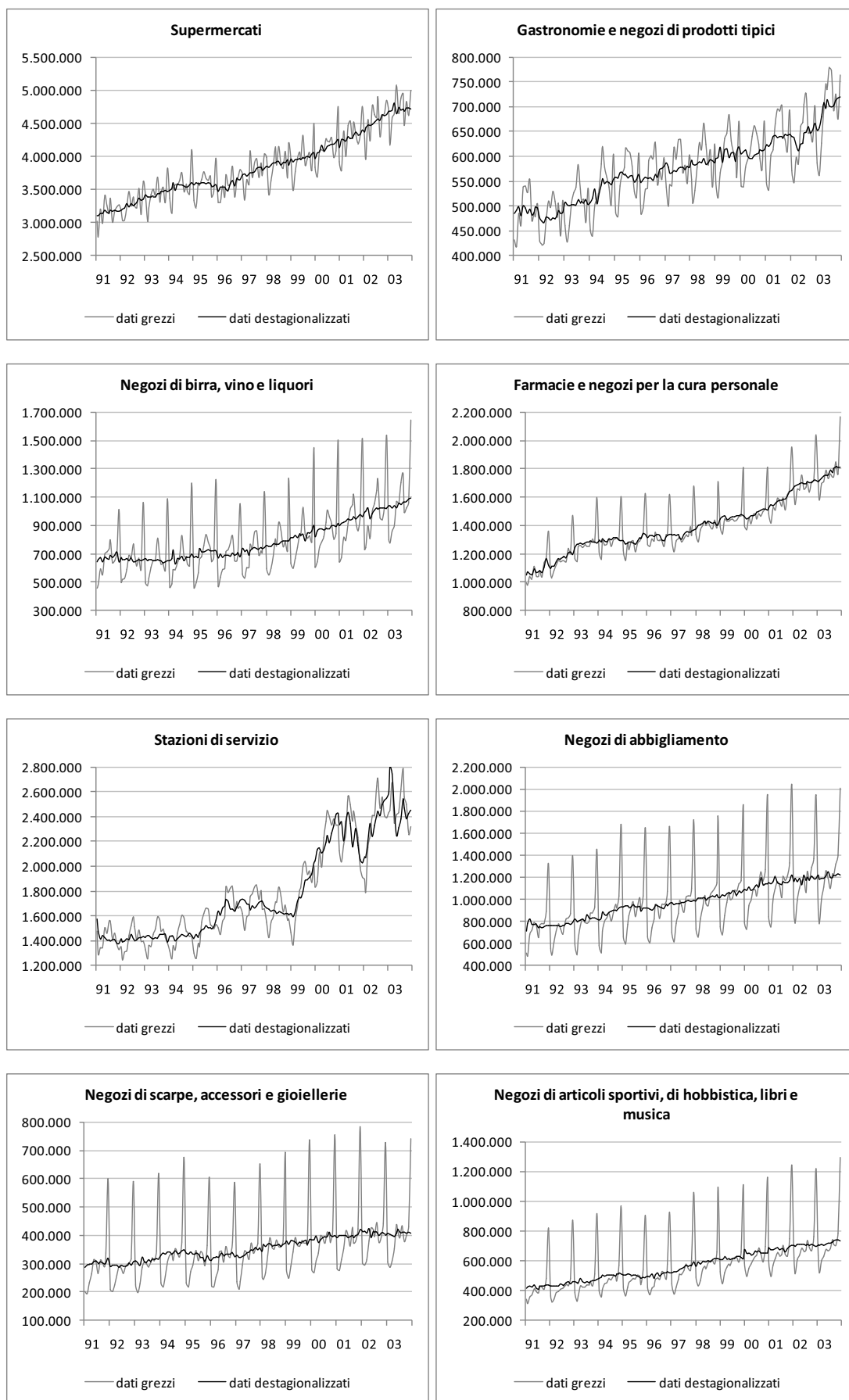


Figura 5.8 Vendite mensili al dettaglio (serie grezze e serie destagionalizzate) da gennaio 1991 a dicembre 2003 nelle province e nei territori canadesi. Prima del 1999 il territorio del Nunavut (PR62) faceva parte dei Territori del Nord-Ovest.

Venditori di auto nuove**Venditori di auto usate e di pezzi di ricambio****Negozi di mobili****Negozi di articoli per la casa****Negozi di computer e software****Negozi di elettrodomestici ed elettronica****Ferramenta****Negozi di materiale per l'edilizia ed il giardinaggio**



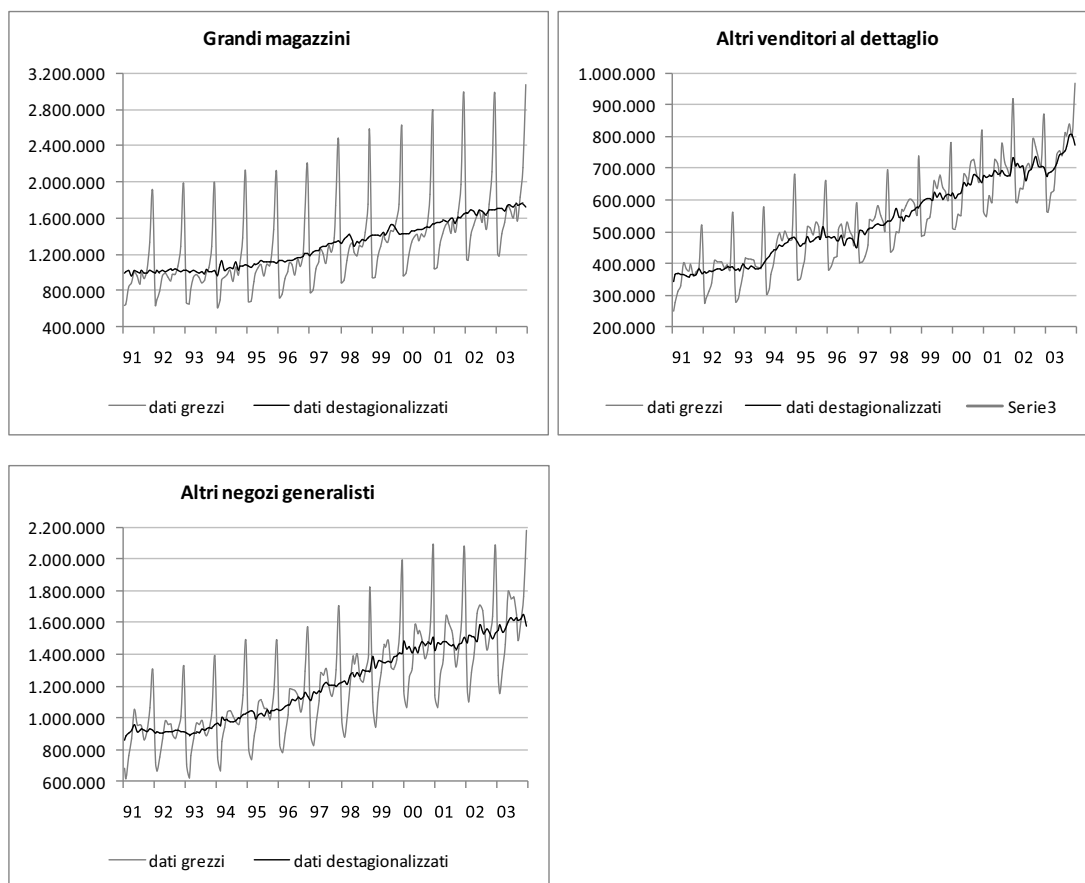


Figura 5.9 Vendite mensili al dettaglio (serie grezze e serie destagionalizzate) da gennaio 1991 a dicembre 2003 per settori commerciali.

Come si nota dai grafici, i dati relativi al territorio del Nunavut sono disponibili da gennaio 1999, data di istituzione dello stesso. Poiché prima di quella data il Nunavut faceva parte dei Territori del Nord-Ovest, si spiega anche lo scalino che nello stesso anno si può vedere nel grafico dei Territori del Nord-Ovest.

In una logica di sistema, volendo rappresentare le serie in forma matriciale, ciò comporta che la colonna relativa al territorio del Nunavut presenti degli zeri in corrispondenza dei primi 8 anni (96 mesi). Ovviamente questi zeri sono strutturali, quindi devono rimanere tali anche al termine della procedura di bilanciamento. La presenza di anni in cui tutti i valori mensili sono nulli, deve essere tenuta in considerazione, quando si opera con tecniche di riconciliazione simultanea, nel definire sia la funzione obiettivo che i vincoli di aggregazione.

Come detto, le serie destagionalizzate sono ottenute da Statistics Canada usando la procedura X11-ARIMA. Le procedure di destagionalizzazione producono due effetti indesiderati:

- (i) i totali annuali della serie destagionalizzata non corrispondono a quelli della serie grezza;
- (ii) le somme contemporanee mensili delle serie destagionalizzate dei due sistemi non coincidono tra loro (e, verosimilmente, nemmeno con la serie dei totali, destagionalizzata per via diretta, che però non è stata fornita da Statistics Canada).

Questi effetti non inficiano la validità della procedura, ma danno luogo a tavole che non soddisfano i vincoli. Gli utenti, invece, preferiscono disporre di dati in cui vengono soddisfatte le relazioni contabili naturali, cioè i vincoli temporali e la corrispondenza tra la somma dei dati disaggregati e il totale Canada.

Per fissare le idee, nel seguito i dati mensili grezzi saranno utilizzati come benchmark annuale (zt); si dispone inoltre delle serie disaggregate dei dati destagionalizzati ma non, come detto, del totale canadese destagionalizzato per via diretta.

Il problema è quindi quello di riconciliare i dati destagionalizzati affinché soddisfino i vincoli di cui sopra. In pratica, riproduciamo la stessa operazione affrontata dall'istituto statistico canadese prima della pubblicazione dei dati MRTS destagionalizzati. La stessa applicazione è stata proposta in letteratura da Di Fonzo e Marini (2003, 2007a), che presentano l'estensione multivariata di Denton, e da Quenneville e Rancourt (2005), che presentano la procedura basata sul modello di regressione lineare illustrata nel paragrafo 4.5.

Le elaborazioni sono state effettuate con Matlab® 7.7.0.471 (R2008b).

I tempi di elaborazione, quando presenti, sono ottenuti con un PC con processore AMD Athlon™ 64 X2 Dual Core Processor 5600+, 2.84Ghz, 3Gb RAM, con Windows™ XP.

5.5 La riconciliazione del sistema MRTS per provincia/territorio con un vincolo esogeno

5.5.1 Vincoli temporali e benchmark univariato

Il primo passo consiste nel benchmark univariato di ciascuna singola serie di ciascun sistema.

Il metodo utilizzato è la variante esatta del metodo *Proportional First Difference* proposto da Denton e presentato a pagina 33. La scelta di un metodo proporzionale appare obbligata, alla luce delle differenze di peso tra provincia e provincia (Tabella 5.2): i tre territori, infatti, contribuiscono complessivamente allo 0,5% del totale della popolazione mentre le due province più abitate e più note (Ontario e Quebec) costituiscono il 60% delle vendite al dettaglio canadesi; quattro province (aggiungendo alle due precedenti Columbia Britannica e Alberta) formano l'85% del totale canadese. Lo stesso vale per la suddivisione in settori commerciali (Tabella 5.3): i supermarket ad esempio incidono per il 20%, mentre i negozi di computer per meno dell'1%.

Il bilanciamento viene ottenuto mediante l'uso di uno script Matlab® che applica la formula (3.2). Il programma è riportato nell'appendice C.4.

La Tabella 5.4 riporta le differenze tra le somme annuali delle serie destagionalizzate e i totali annui delle serie grezze usate come benchmark.

	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
PR10	1	1	0	1	0	0	0	1	0	1	0	0	0
PR11	1	0	0	0	1	1	1	1	0	1	0	0	1
PR12	0	-2	1	0	0	1	0	1	0	-1	1	2	0
PR13	-2	0	0	-1	0	0	1	1	1	1	0	0	0
PR24	1	5	2	9	11	11	7	6	9	-2	-4	6	6
PR35	2	4	2	3	3	2	3	2	-1	-1	0	4	21
PR46	1	0	0	-1	1	-1	1	1	-2	0	0	1	0
PR47	1	0	0	0	2	1	0	0	-1	0	0	0	1
PR48	2	0	1	1	0	1	2	2	1	1	1	3	1
PR59	2	1	1	2	3	2	2	2	3	0	2	1	1
PR60	2	2	0	2	0	1	-1	1	2	0	-1	0	-1
PR61	-1	-1	0	-1	-1	0	1	0	0	2	3	-2	1
PR62	*	*	*	*	*	*	*	*	0	1	2	0	1

Tabella 5.4 Discrepanze tra i totali annuali destagionalizzati e grezzi, sistema PR (milioni di dollari).

(*) Prima del 1999 il territorio del Nunavut (PR62) non esisteva.



Figura 5.10 Discrepanze tra i totali annuali destagionalizzati e grezzi per le province che presentano le differenze maggiori (milioni di dollari).

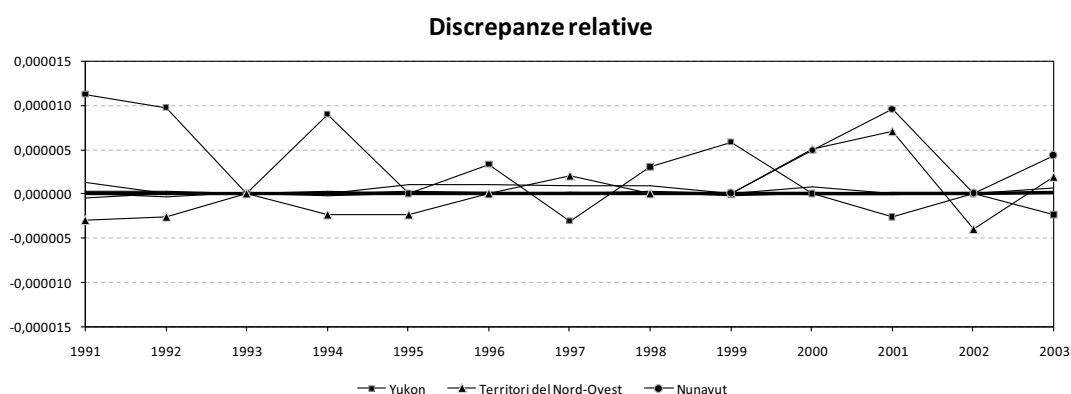


Figura 5.11 Discrepanze relative tra i totali annuali destagionalizzati e grezzi per le tredici province/territori canadesi. In evidenza le serie che presentano le discrepanze maggiori.

In Figura 5.10 vengono mostrate le due regioni che presentano le maggiori differenze (in milioni di dollari canadesi): Quebec e Ontario. Si può notare come solitamente la differenza sia positiva e raramente superi i 10 milioni, anche se nel 2003 in Onta-

rio la discrepanza è di 21 milioni di dollari. In generale comunque le differenze sono praticamente irrilevanti, come si può vedere dal grafico delle discrepanze relative (Figura 5.13), e non sarà dedicata ulteriore attenzione alla procedura di benchmark univariato. Questa, pur non presentando particolari spunti di interesse teorico, è di estrema importanza pratica, costituendo il primo passo della procedura a due step che verrà proposta in seguito.

5.5.2 Vincolo di aggregazione contemporanea esogeno

Nella procedura di destagionalizzazione, la serie destagionalizzata per l'intera economia canadese si può ottenere in due modi:

- (i) per via diretta, sottoponendo alla procedura di destagionalizzazione non solo le m serie storiche dei dati suddivisi per provincia/territorio ma anche la serie grezza del totale, o
- (ii) per via indiretta, considerando come serie destagionalizzata la somma delle serie destagionalizzate disaggregate.

In questo esempio, considereremo come benchmark contemporaneo il totale ottenuto sommando le serie destagionalizzate della tavola TG⁶, e lo applicheremo al sistema PR. Questo tipo di vincolo è detto esogeno, perché è esterno al sistema che si sta considerando.

La Figura 5.13 mostra le differenze mensili, in livello e percentuali, tra la somma delle 13 serie regionali (province e territori) e il benchmark.

Queste differenze variano, generalmente, tra il -1% e l'1, con due sole "punte" intorno all'1,5% in valore assoluto, e fluttuano attorno allo zero. Il correlogramma empirico della serie delle discrepanze (Figura 5.12) evidenzia l'esistenza di un pattern di dipendenza temporale, confermato dal test di Ljung-Box ($Q(39) = 146,96$; $p\text{-value} = 0$) che rifiuta l'ipotesi nulla. Potrebbe essere utile capire se da ciò può scaturire qualche informazione utile per una procedura di riconciliazione, ma

⁶ In effetti il benchmarking contemporaneo è ottenuto con l'applicazione di una procedura di benchmark univariato alla somma delle serie disaggregate del sistema TG, come approfondito nell'appendice B.3.

questi approfondimenti esulano dagli obiettivi principali di questa tesi e la questione viene pertanto lasciata aperta.

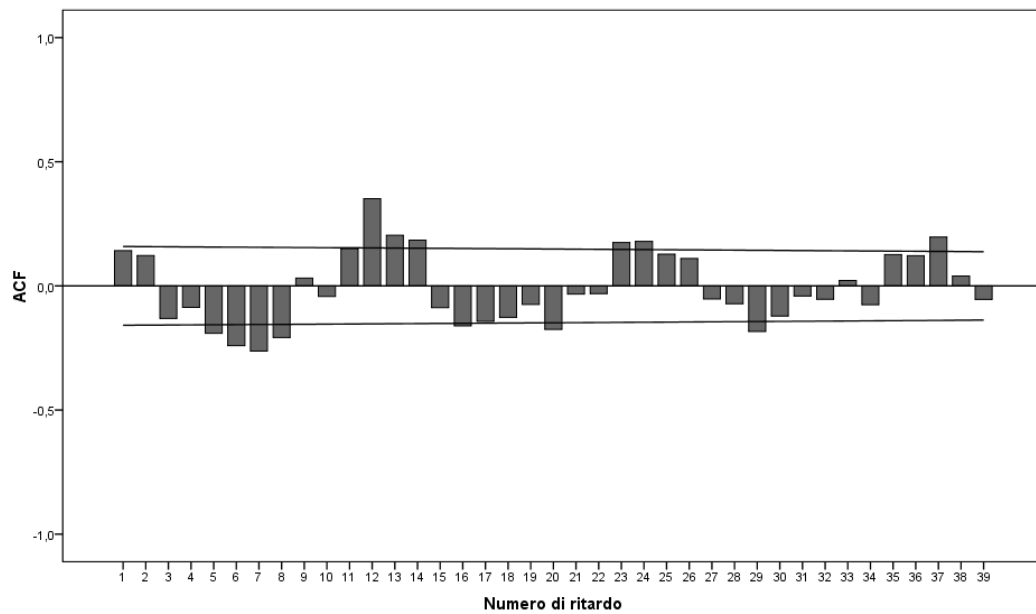


Figura 5.12 Autocorrelogramma empirico della serie delle discrepanze, con bande di confidenza nell'ipotesi di processo white noise.

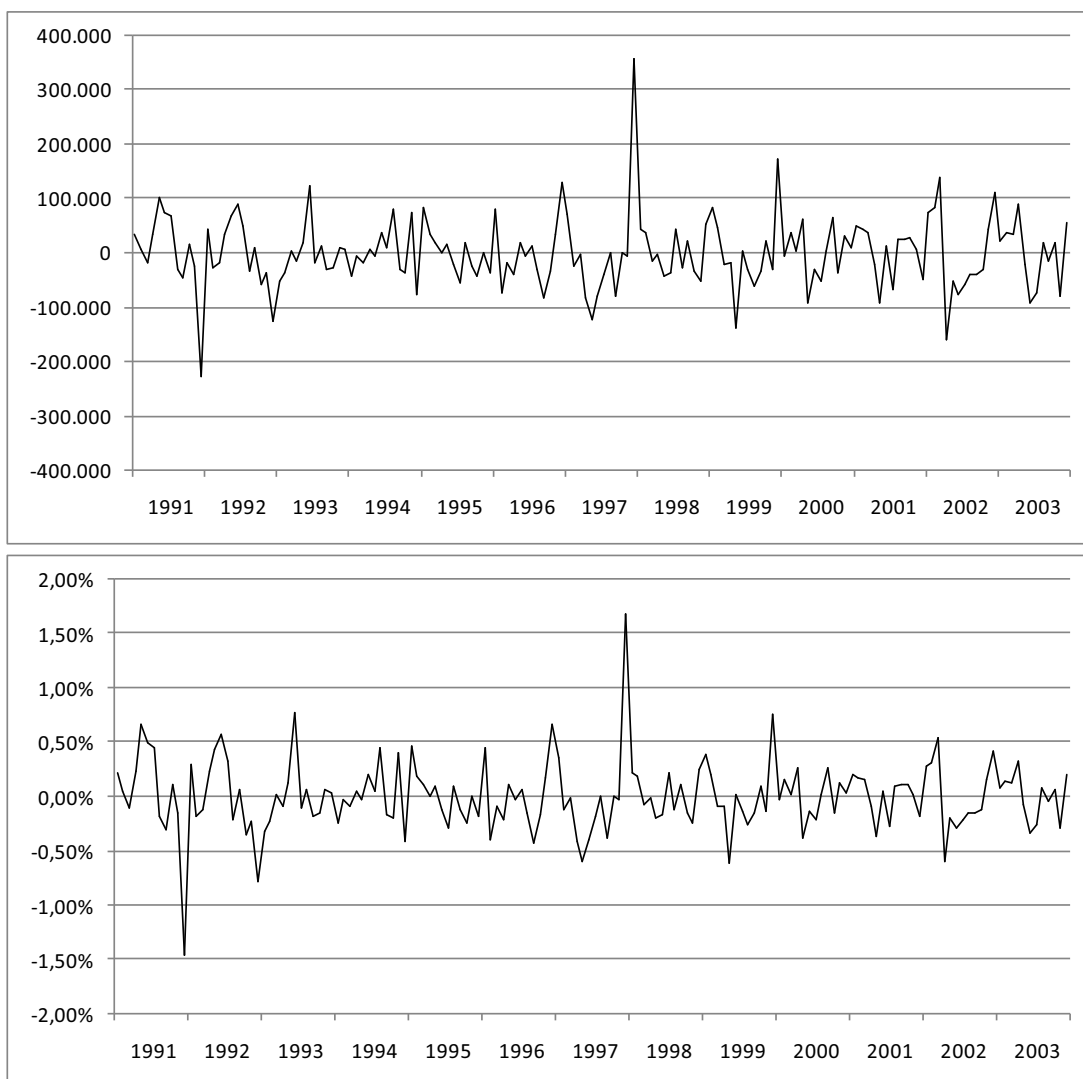


Figura 5.13 Differenze mensili (in milioni di dollari, sopra, e percentuali, sotto) tra la somma delle 13 serie per provincia/territorio destagionalizzate e il totale nazionale destagionalizzato, ricavato dal sistema per settori commerciali.

5.5.3 Procedure di riconciliazione e valutazione dei risultati

Le procedure di riconciliazione che saranno analizzate sono:

- (1) l'estensione multivariata di Denton (si veda paragrafo 4.1), nella variante anno-per-anno (DENT1). La procedura è proposta anche sulle serie preliminari già riallineate alle rispettive serie annuali con Denton PFD (DENT1*).
- (2) L'estensione multivariata simultanea di Denton, proposta sia per le serie preliminari originarie (DENTF) che per le serie riallineate singolarmente con Denton PFD (DENTF*).

- (3) La procedura di Stone (STONE, paragrafo 2.4), che non considera la dinamica temporale ma si occupa di bilanciare anno-per-anno la matrice delle serie mensili ai totali di riga e di colonna.
- (4) La procedura di Quenneville e Rancourt (QR, paragrafo 4.5), che preserva la dinamica temporale applicando Denton univariato al primo step, prima di effettuare il bilanciamento secondo Stone.

La bontà di ciascuna procedura di riconciliazione viene misurata attraverso una opportuna sintesi della correzione apportata dai valori riconciliati ai dati. Con riferimento ai tassi di crescita, l'aggiustamento è dato da:

$$d_{t,j} = \begin{cases} \left[\left(\frac{y_{t,j} - y_{t-1,j}}{y_{t-1,j}} \right) - \left(\frac{p_{t,j} - p_{t-1,j}}{p_{t-1,j}} \right) \right] \cdot 100 & t = 2, \dots, n \quad \text{se } y_{t-1,j} \neq 0 \\ 0 & \text{se } y_{t-1,j} = p_{t-1,j} = 0 \end{cases} \quad (5.1)$$

Un indice sintetico dell'aggiustamento apportato a ciascuna serie è dato dalla radice quadrata dell'aggiustamento quadratico medio (RMSA, *Root Mean Squared Adjustment*), definito come

$$RMSA_j = \sqrt{\frac{1}{n_{NZ_j} - 1} \sum_{t=2}^n d_{t,j}^2} \quad j = 1, \dots, m \quad (5.2)$$

con n_{NZ_j} numero di elementi diversi da 0 nella j -esima serie storica. Ad esempio, nel caso del sistema PR, $n_{NZ_j} = 156$ per $j = 1, \dots, 12$ $n_{NZ_j} = 60$ per $j = 13$.

Una sintesi per l'intero sistema è facilmente ottenibile come

$$RMSA = \sqrt{\frac{1}{\sum_{j=1}^m (n_{NZ_j} - 1)} \sum_{j=1}^m (n_{NZ_j} - 1) RMSA_j^2}. \quad (5.3)$$

L'RMSA complessivo, come si capisce facilmente dalla formula, è una media⁷ ponderata degli $RMSA_j$ delle singole serie. Si noti che questa media assegna uguale importanza a ciascuna serie, mentre, come abbiamo fatto precedentemente notare, sia

⁷ Si tratta di una media di potenza di ordine 2.

nel caso della classificazione per province sia in quello della classificazione per settori commerciali, l'importanza tra le serie è notevolmente diversa.

Per questo motivo può risultare più utile un indicatore pesato, che tenga conto dell'importanza di ogni singola serie in ciascun istante temporale, chiamato WRMSA (*Weighted Root Mean Squared Adjustment*). Indicato con $w_{t,j}$ il sistema dei pesi,

$w_{t,j} = \frac{y_{t-1,j}}{\sum_{j=1}^m y_{t-1,j}}$, $w_{t,j} \geq 0$, $\sum_{j=1}^m w_{t,j} = 1$, l'indice è definito come

$$WRMSA = \sqrt{\frac{1}{n-1} \sum_{j=1}^m \sum_{t=2}^n w_{t,j} d_{t,j}^2}. \quad (5.4)$$

Un secondo indice della bontà di adattamento ai dati è dato dalla funzione criterio vista in (4.14). Ovviamente, per definizione, la procedura simultanea che generalizza il criterio univariato di Denton genera valori minimi per quest'ultimo criterio: questo ci permette di valutare di quanto la procedura di Quenneville e Rancourt se ne discosta.

5.5.4 Esito e confronto dei risultati

La Tabella 5.5 mostra il valore dell'RMSA per ciascuna singola serie, dell'RMSA complessivo e del tempo computazionale per ciascuna delle procedure di riconciliazione

	DENT1	DENT1*	DENTF	DENTF*	STONE	QR
PR10	0,0312	0,0312	0,0312	0,0312	0,4234	0,4234
PR11	0,0078	0,0078	0,0078	0,0078	0,4236	0,4236
PR12	0,0598	0,0598	0,0591	0,0591	0,4249	0,4249
PR13	0,0461	0,0461	0,0458	0,0458	0,4220	0,4220
PR24	0,4158	0,4158	0,4166	0,4166	0,4170	0,4170
PR35	0,6912	0,6912	0,6903	0,6903	0,4223	0,4223
PR46	0,0610	0,0610	0,0610	0,0610	0,4219	0,4219
PR47	0,0545	0,0545	0,0546	0,0546	0,4222	0,4222
PR48	0,1974	0,1974	0,1963	0,1963	0,4236	0,4236
PR59	0,2541	0,2541	0,2558	0,2558	0,4200	0,4200
PR60	0,0022	0,0022	0,0022	0,0022	0,4253	0,4252
PR61	0,0034	0,0034	0,0034	0,0034	0,4242	0,4242
PR62	0,0010	0,0010	0,0010	0,0010	0,3401	0,3400
Intero sistema (RMSA)	0,2490	0,2490	0,2490	0,2490	0,4202	0,4202
Intero sistema (WRMSA)	0,4837	0,4837	0,4835	0,4835	0,4210	0,4210
Tempo computazionale	0,0337	0,0336	0,0220	0,1573	0,0175	0,0270

Tabella 5.5 RMSA delle serie riconciliate con diverse procedure e relativi tempi di calcolo. L'asterisco indica le procedure che prevedono l'uso della tecnica Denton PFD univariato al primo step (QR=STONE*).

La prima, importante, nota riguarda i tempi computazionali, che sono sempre molto piccoli: 2/3 centesimi di secondo. Questo aspetto è molto importante: prima che fossero disponibili i calcolatori oggi in commercio alcune delle procedure analizzate nei primi capitoli di questa tesi non potevano essere applicate o, quando ciò era possibile, i tempi di attesa potevano essere abbastanza elevati. Un ulteriore contributo al miglioramento dei tempi di elaborazione è dovuto agli algoritmi di gestione delle matrici sparse di Matlab, in particolare dalla versione 7.5 (commercialmente nota come r2007b) in poi.

La statistica RMSA, graficamente illustrata in Figura 5.14, permette invece alcune considerazioni: la procedura di Stone, come ci si poteva aspettare, è quella che modifica maggiormente i tassi di crescita; i risultati migliori, e anche questo è un risultato ampiamente previsto, si ottengono dall'estensione multivariata della tecnica di Denton (DENTF).

Si fa notare come, a causa delle esigue discrepanze temporali, il primo step con Denton univariato non modifica, di fatto, i dati e quindi le procedure a due step presentano risultati identici (DENT1 con DENT1* e DENTF con DENTF*) o con differenze irrilevanti (STONE e QR=STONE*).

L'ultima osservazione sull'RMSA riguarda i valori dell'RMSA di ciascuna variabile: i tassi di crescita maggiormente modificati dalla riconciliazione con l'estensione multivariata di Denton sono quelli relativi alle serie con valori più grandi: come si può vedere dalla Figura 5.15 vengono infatti modificati maggiormente i dati delle province più popolate, il che è in qualche modo controintuitivo, poiché gli esperti di Statistics Canada considerano proprio queste serie come più affidabili.

Per comprendere meglio questo aspetto, consideriamo la statistica WRMSA, che, come detto, si ottiene ponderando l'RMSA di ciascuna serie per il peso della serie stessa, dato dalle quote di mercato medie di ciascuna regione.

Secondo questo criterio, come si può vedere anche dalla Figura 5.14, la procedura di Quenneville e Rancourt è quella che fornisce i risultati migliori, perché preserva

maggiormente le serie relative alle provincie più abitate, che sono considerate migliori.

La questione è tuttavia piuttosto delicata, e la precedente conclusione lascia spazio a qualche perplessità: si ritiene, infatti, che non debba essere lo strumento, ovvero la procedura di riconciliazione, a stabilire *tout court* che le serie numericamente più importanti siano le più affidabili, quanto piuttosto riscontri empirici derivanti, ad esempio, dalla procedura statistica di stima delle serie grezze e/o dagli esiti dell'applicazione delle tecniche di destagionalizzazione (ad esempio, standard error delle serie destagionalizzate), o -ancora- da opinioni di esperti. Questa informazione, sia pure talvolta soggettiva, può essere usata in forma di coefficienti di alterabilità, sia nella procedura di Quenneville e Rancourt, che, soprattutto, nell'estensione multivariata di Denton.

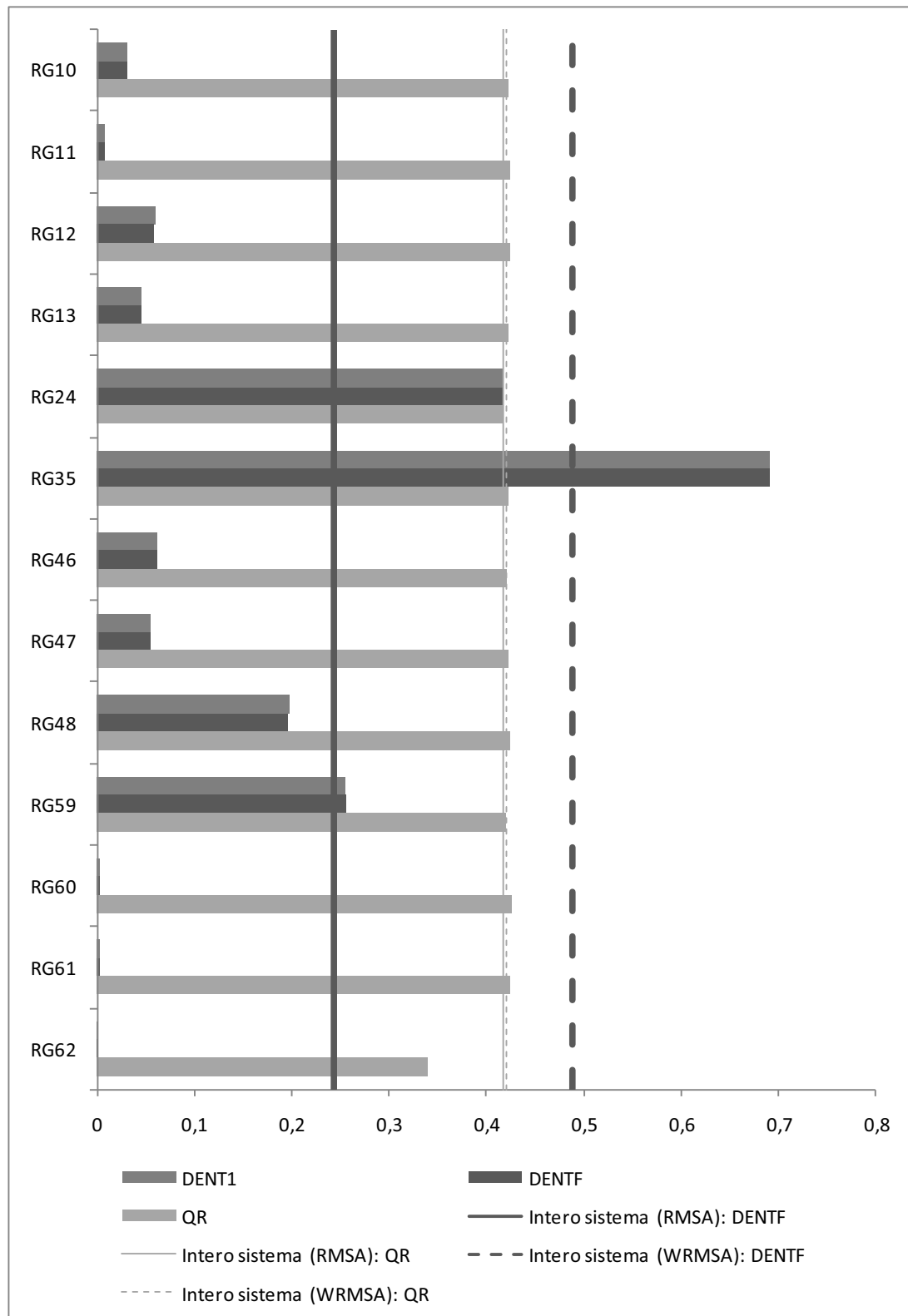


Figura 5.14 Confronto tra i diversi RMSA.

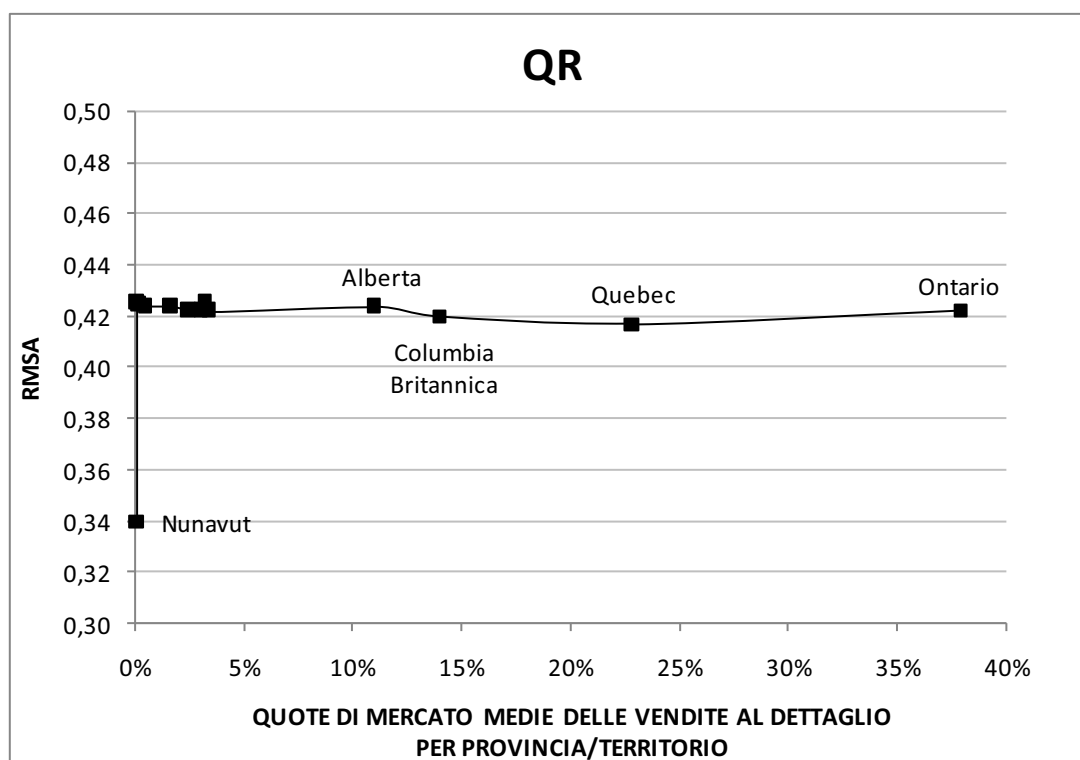
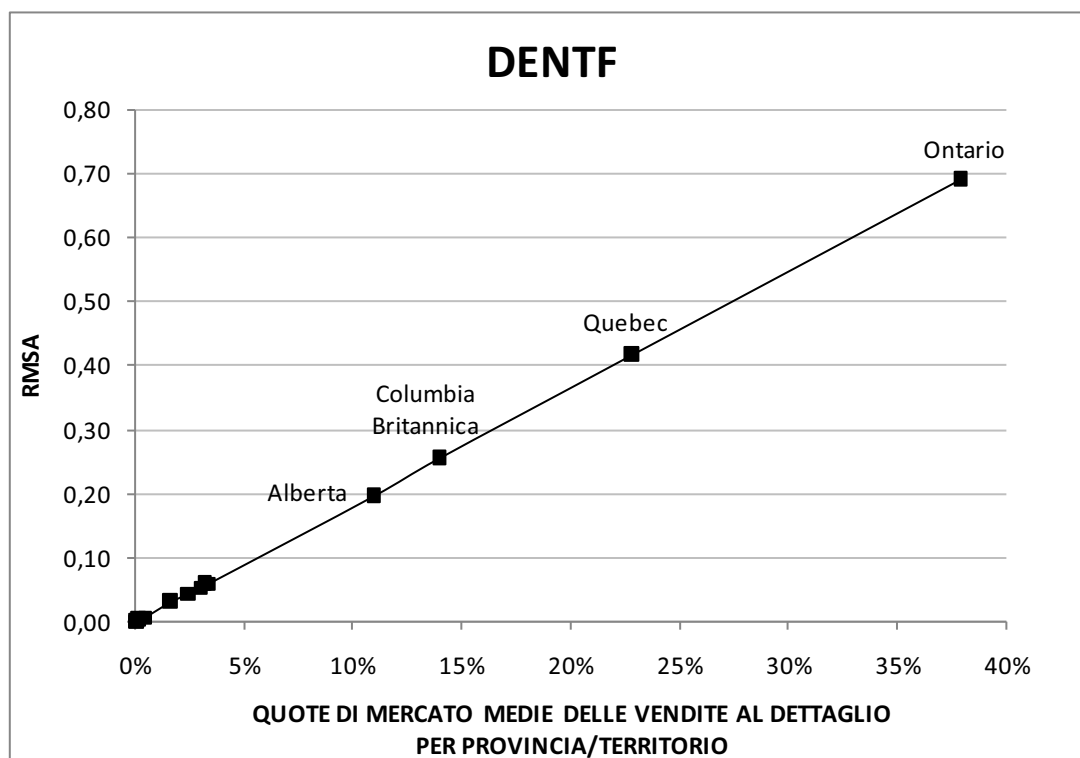


Figura 5.15 Grafici di dispersione tra le vendite medie al dettaglio (in milioni di dollari) per province e territori e la statistica RMSA secondo il procedimento di Quenneville e Rancourt (QR) e l'estensione multivariata simultanea di Denton (DENTF).

Per concludere la Tabella 5.6 riporta i valori della statistica PFD in corrispondenza delle diverse procedure di riconciliazione considerate.

	DENT1	DENT1*	DENTF	DENTF*	STONE	QR
PR10	0,000015	0,000015	0,000015	0,000015	0,002737	0,002737
PR11	0,000001	0,000001	0,000001	0,000001	0,002735	0,002735
PR12	0,000054	0,000054	0,000053	0,000053	0,002735	0,002735
PR13	0,000033	0,000033	0,000032	0,000032	0,002736	0,002736
PR24	0,002718	0,002718	0,002726	0,002726	0,002735	0,002735
PR35	0,007332	0,007332	0,007311	0,007311	0,002735	0,002735
PR46	0,000057	0,000057	0,000057	0,000057	0,002737	0,002737
PR47	0,000046	0,000046	0,000046	0,000046	0,002735	0,002735
PR48	0,000594	0,000594	0,000587	0,000587	0,002736	0,002736
PR59	0,001002	0,001002	0,001015	0,001015	0,002738	0,002738
PR60	0,000000	0,000000	0,000000	0,000000	0,002735	0,002735
PR61	0,000000	0,000000	0,000000	0,000000	0,002733	0,002733
PR62	0,000000	0,000000	0,000000	0,000000	0,000678	0,000677
Intero sistema	0,011851	0,011851	0,011844	0,011844	0,033506	0,033505

Tabella 5.6 Criterio PFD per diversi metodi di riconciliazione. L'asterisco indica le procedure che prevedono l'uso della tecnica Denton PFD univariato al primo step (QR=STONE*).

La prima osservazione ci porta a ripetere che, a causa delle discrepanze temporali praticamente irrilevanti, il primo step non provoca nessun effetto pratico, e i valori di DENT1 e DENT1* e di DENTF e DENTF* sono identici, mentre quelli di STONE e di QR (=STONE*) presentano differenze di scarsa entità.

Com'è ovvio, DENTF fa registrare il PFD di sistema più basso. DENT1 si attesta su valori molto simili a quelli di DENTF, ad ulteriore conferma del fatto che, nel caso in questione, le due procedure portano a risultati molto simili: vanno invece rimarcate le differenze a livello di singola serie tra la procedura QR e quelle fondate sul principio simultaneo di Denton. Anche il criterio PFD, infatti, evidenzia come nel caso di DENT1 e DENTF le serie con valori più grandi siano maggiormente interessate dall'aggiustamento; la procedura QR, invece, effettua una ripartizione delle discrepanze temporali sostanzialmente equidistribuita, con -ancora una volta- la solita eccezione del territorio del Nunavut (PR62), dove il valore più basso non può essere spiegato soltanto dalla minore lunghezza della serie.

5.6 La riconciliazione del sistema per settore commerciale con un vincolo esogeno

Avendo deciso di riconciliare i dati usando il totale per settori commerciali quale benchmark esogeno, le discrepanze contemporanee (Figura 5.16) sono molto esigue, così come, per i motivi già illustrati, quelle temporali.

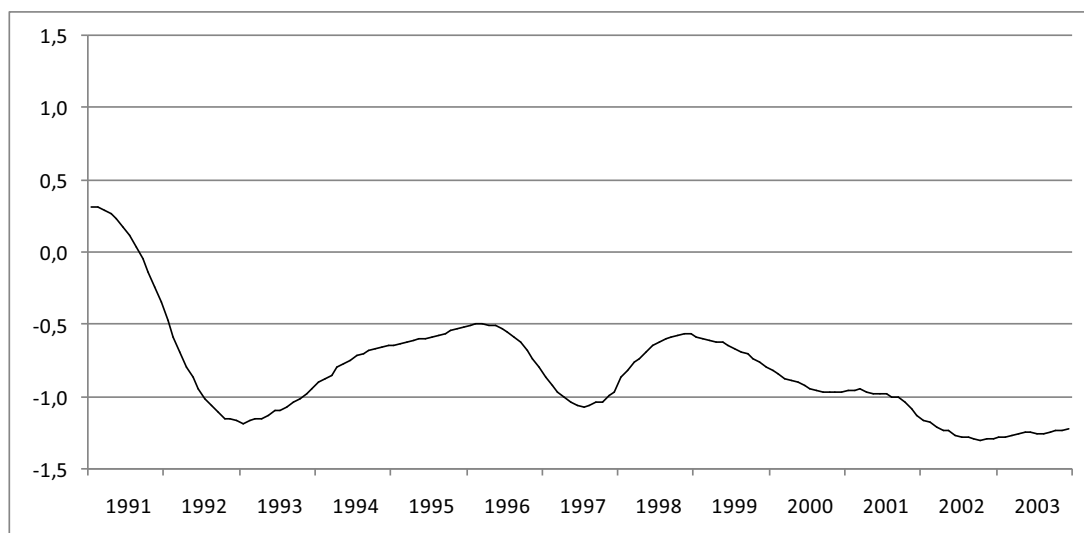


Figura 5.16 Differenze mensili (in milioni di dollari) tra la somma delle 19 serie per settori commerciali destagionalizzate e il totale nazionale.

Per questo, l'applicazione non ha nessun interesse teorico, ma in assenza della serie Totale Canada destagionalizzata per via diretta, questo modo di procedere è l'unica via ragionevole basata su un vincolo esogeno e completa la riconciliazione dei due sistemi.

Le esigue discrepanze mostrano RMSA praticamente nulli per tutte le tecniche considerate, mentre più interessante, perché ci permetterà confronti con le procedure basate su vincoli diversi, è la statistica RMSA di entrambi i sistemi. Dopo aver riconciliato singolarmente ciascuno dei due sistemi, si è formata un'unica matrice con i dati riconciliati e un'unica matrice di dati preliminari ed è stato calcolato l'RMSA.

	DENTF	QR
RMSA	0,1564	0,2640
WRMSA	0,3421	0,2977

Tabella 5.7 RMSA e WRMSA per DENTF e QR.

5.7 La riconciliazione in presenza di discrepanze temporali

L'esercizio mostrato finora, basato su veri valori, non permetteva un confronto completo tra le procedure; in particolare non esistono differenze rilevanti tra le procedure descritte e le analoghe metodologie a due step, a causa delle esigue discrepanze temporali.

Si è quindi svolto l'esercizio modificando i valori come segue:

- le serie preliminari ad alta frequenza, p , e la serie dei vincoli di aggregazione contemporanea, zc , sono state conservate;
- le serie a bassa frequenza, zt , sono state invece sostituite con altre ottenute con l'ausilio della generazione di numero pseudo-casuali per mezzo dell'elaboratore elettronico. Non trattandosi di un vero e proprio esercizio di simulazione, non si è prestata particolare attenzione nella procedura di randomizzazione, descritta nell'appendice B.4.

La Figura 5.17 mostra le differenze percentuali tra la somma annuale dei valori preliminari p e il relativo benchmark, zt .

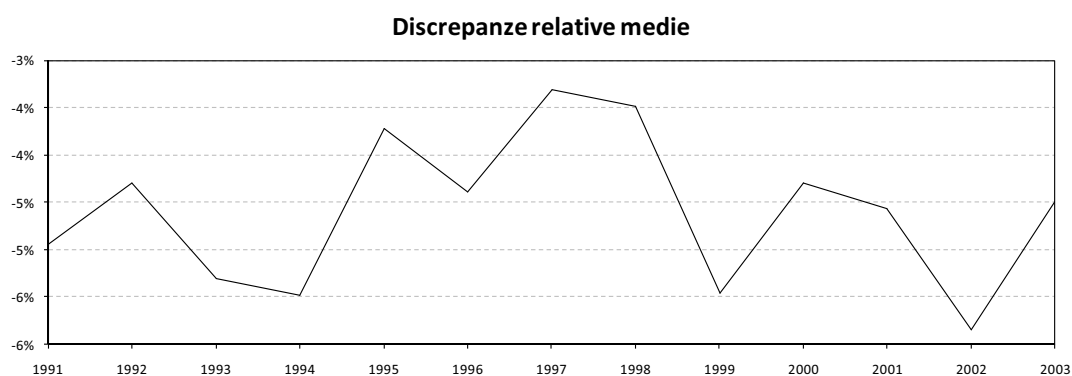


Figura 5.17 Discrepanze relative medie tra i totali annuali destagionalizzati e grezzi per le tredici province/territori canadesi.

	DENT1	DENT1*	DENTF	DENTF*	STONE	QR
PR10	0,8409	0,3496	0,3496	0,3496	0,9420	0,5516
PR11	1,1707	0,4980	0,4980	0,4980	1,2787	0,6506
PR12	0,7358	0,3020	0,3014	0,3018	0,8148	0,5207
PR13	0,8093	0,3559	0,3558	0,3559	1,0701	0,5496
PR24	0,9608	0,5489	0,5417	0,5495	0,9697	0,5441
PR35	1,0312	0,8422	0,8518	0,8412	0,9382	0,6417
PR46	0,8352	0,3295	0,3293	0,3295	0,9203	0,5330
PR47	1,3223	0,5506	0,5505	0,5506	1,6539	0,6954
PR48	0,9071	0,4452	0,4425	0,4447	0,9435	0,5804
PR59	1,2298	0,5941	0,5908	0,5948	1,2840	0,6790
PR60	1,2236	0,5343	0,5343	0,5343	1,3701	0,6862
PR61	1,0208	0,3945	0,3944	0,3944	1,0632	0,5821
PR62	1,0881	0,3912	0,3912	0,3912	1,0579	0,5164
Intero sistema (RMSA)	1,0260	0,4972	0,4973	0,4972	1,1264	0,6019
Intero sistema (WRMSA)	1,0217	0,6559	0,6583	0,6557	1,0326	0,6116

Tabella 5.8 RMSA delle serie riconciliate con diverse procedure. L'asterisco indica le procedure che prevedono l'uso della tecnica Denton PFD univariato al primo step (QR=STONE*).

Rispetto a quanto già detto sinora la Tabella 5.9 ci permette di far notare come, in presenza di discrepanze temporali, le procedure a due step forniscano risultati migliori, seppur con differenze tra le diverse procedure.

Nel caso dell'estensione multivariata di Denton applicata anno-per-anno (DENT1) le differenze sono dovute essenzialmente agli step che si vengono a creare tra l'ultimo periodo di un anno e il primo periodo di quello successivo. Questo non accade nel caso dell'applicazione simultanea di Denton (DENTF), dove la dinamica temporale viene preservata, per costruzione, dalla procedura stessa e le differenze sono irrilevanti (la statistica RMSA vale 0,493 in entrambi i casi).

Le cose cambiano per quanto riguarda la procedura di Stone che è, in effetti, una procedura per il bilanciamento di tavole di aggregati che non preserva la dinamica temporale. Questo è confermato dal valore dell'indice RMSA e WRMSA, che sono più alti. L'applicazione del primo step, proposta da Quenneville e Rancourt, permette invece di ottenere la riconciliazione di tavole di serie storiche in maniera semplice, essendo la dinamica temporale preservata proprio da Denton univariato.

5.8 La riconciliazione dei sistemi per provincia/territorio e per settore commerciale con un vincolo endogeno

5.8.1 La scelta del vincolo di aggregazione contemporanea

Se la scelta del vincolo di aggregazione temporale di dati destagionalizzati è in qualche modo naturale, non così quella del vincolo contemporaneo. Tra i vincoli che possono essere ragionevolmente scelti abbiamo:

- vincoli esogeni:
 - serie di valori che vengono considerati noti e, dunque, non appartenenti al sistema;
 - ottenuti dalla destagionalizzazione diretta della serie grezza Totale Canada,
- vincoli endogeni: ottenuti ponendo che la somma degli aggregati di un sistema sia uguale alla somma degli aggregati del secondo sistema.

Poniamo ora l'attenzione all'ultimo tipo di vincolo. Vogliamo cioè che i dati soddisfino la relazione

$$\sum_{j=1}^{13} y_j^{PR} = \sum_{k=1}^{19} y_k^{TG} \quad (5.5)$$

Usando una notazione matriciale, possiamo creare una matrice Y_i , 156x32, dove le prime dodici colonne hanno le serie mensili, per un totale di 13 anni, relative alle prime dodici provincie e territori canadesi, la tredicesima, relativa al territorio del Nunavut che esiste a partire dal 1999, è nulla nelle prime 96 righe, e le successive 19 sono le serie mensili per settore commerciale:

$$Y_{(156 \times 32)} = \begin{bmatrix} y_1^{PR} & y_2^{PR} & \dots & y_{13}^{PR} & y_1^{TG} & y_2^{TG} & \dots & y_{19}^{TG} \\ y_1 & y_2 & \dots & y_{13} & y_{14} & y_{15} & \dots & y_{32} \end{bmatrix}.$$

La (5.5) diventa:

$$\sum_{j=1}^{13} y_j = \sum_{j=20}^{32} y_j. \quad (5.6)$$

Ulteriori approfondimenti sulla costruzione dei vincoli e della funzione obiettivo sono riportati nell'appendice B.5.

5.8.2 Valutazione delle performance

Nel proseguo si porrà attenzione alle due procedure che danno i risultati migliori: DENTF, cioè la riconciliazione secondo l'estensione multivariata simultanea di Denton, e QR, la procedura a due step introdotta da Quenneville e Rancourt.

Le conclusioni non differiscono da quelle a cui si è giunti nel caso di un vincolo esogeno:

- la procedura DENTF ottiene risultati mediamente migliori (Tabella 5.9), modificando però molto di più gli aggregati numericamente più importanti;
- la procedura QR presenta risultati tutti molto vicini alla media, introducendo quindi discrepanze tra i tassi di crescita più simili. Infatti, gli RMSA ponderati (Tabella 5.10) premiano, ancora una volta, questa procedura.

RMSA	DENTF	QR
Vincolo esogeno	0,1564	0,2640
Vincolo endogeno	0,1294	0,2107

Tabella 5.9 Confronto tra gli RMSA ottenuti nel sistema con vincolo esogeno (riconciliazione ad una via) e nel sistema con vincolo endogeno (riconciliazione a due vie marginale).

WRMSA	DENTF	QR
Vincolo esogeno	0,3421	0,2977
Vincolo endogeno	0,2605	0,2105

Tabella 5.10 Confronto tra i WRMSA ottenuti nel sistema con vincolo esogeno (riconciliazione ad una via) e nel sistema con vincolo endogeno (riconciliazione a due vie marginale).

La procedura con vincolo endogeno modifica i tassi di crescita in maniera minore di quanto faccia la riconciliazione ad una via. Questo risultato è atteso, vista la particolare scelta del vincolo di aggregazione contemporanea esogeno: ponendolo uguale alla somma del sistema per settori commerciali quest'ultimo è, di fatto, già riconciliato, viste anche le esigue discrepanze territoriali. Tutte le discrepanze vanno invece distribuite tra gli aggregati del sistema per province/territori, provocando modifiche maggiori ai tassi di crescita del sistema.

Confrontiamo adesso i vincoli implicitamente derivanti dalla riconciliazione marginale a due vie con quello esogeno suggerito dagli esperti di Statistics Canada e assunto come termine di paragone. Per far questo ci limitiamo ad analizzare i tassi di crescita con riferimento al segno assunto dagli stessi, e confrontiamo i punti di svolta.

Parleremo pertanto di corretta classificazione quando il tasso di crescita del Totale Canada ottenuto con vincolo endogeno ha lo stesso segno di quello del Totale Canada dato esogenamente, e di errata classificazione quando ciò non è vero.

Qualitativamente, il profilo dinamico delle serie riconciliate non sembra essere particolarmente influenzato dalla procedura di stima/riconciliazione adottata. Sembra tuttavia esservi una migliore capacità di rispettare il profilo dinamico da parte di DENTF rispetto a QR (Tabella 5.11).

Segni dei tassi di crescita		
Procedura di riconciliazione	Corretta classificazione	Errata classificazione
DENTF	151 (97,42%)	4 (2,58%)
QR	148 (95,48%)	7 (4,52%)

Tabella 5.11 Confronto tra i punti di svolta delle serie Totale Canada data esogenamente e ottenute dalla procedura di riconciliazione con vincolo endogeno.

A conferma della poca influenza della procedura di riconciliazione adottata, facciamo notare che dei 7 tassi di crescita classificati erroneamente dalla procedura QR, 4 sono relativi a mesi in cui anche la procedura DENTF commette un errore di segno.

6 Conclusione

Le tavole di contabilità nazionale ottenute dai produttori di statistiche ufficiali spesso non soddisfano i vincoli di aggregazione, temporali e contemporanei, che l'utente si aspetta a seguito degli schemi contabili usati, e/o della natura delle informazioni e/o delle convenzioni definitorie adottate. Per risolvere questo problema e poter pubblicare statistiche utili devono quindi ricorrere a procedure di bilanciamento. I problemi sorgono quando una delle dimensioni è rappresentata dal tempo: è infatti opportuno che la procedura, che in questo caso chiamiamo di riconciliazione, soddisfi i vincoli preservando, allo stesso tempo, il profilo dinamico delle serie coinvolte.

Per questo la tesi inizia presentando il bilanciamento di matrici dei conti, con le due procedure più diffuse: l'algoritmo di scaling RAS e la procedura ai minimi quadrati, entrambi dovuti a Stone¹. La procedura ai minimi quadrati è di particolare interesse perché costituisce la base della procedura di riconciliazione di Quenneville e Rancourt (2005).

Data una matrice da bilanciare $P = [p_1 \ \cdots \ p_m]$, e indicato con $p = [p'_1 \ \cdots \ p'_m]'$ la sua vettorizzazione, allora il vettore dei valori bilanciati è dato da

$$y = p + VG'(GVG')^{-1}(z - Gp),$$

dove z rappresenta il vettore dei vincoli, V la matrice di varianza-covarianza che Stone ipotizza essere una matrice diagonale, $V = \text{diag}(p)$, e G rappresenta la matrice di aggregazione.

Successivamente ci si è occupati del problema della riconciliazione presentando il caso del benchmarking univariato di una serie storica, basandosi sulla minimizzazione quadratica proposta da Denton (1971). La soluzione è data da

¹ L'algoritmo RAS è stato impiegato, indipendentemente l'uno dall'altro, da diversi autori. A Stone (1962) se ne deve la forma attuale, oltre che il nome.

$$y = p + M^{-1}J'(JM^{-1}J')^{-1}(z_t - Jp),$$

dove z_t è il vettore dei vincoli temporali, la matrice M caratterizza la funzione criterio quadratica che si vuole minimizzare e J è la matrice di aggregazione temporale.

Trattandosi, in entrambi i casi, di problemi di ottimizzazione lineare, non deve stupire che, analiticamente, le due soluzioni sono uguali: sono le forme delle matrici coinvolte e il loro significato in chiave statistica o in relazione al problema che si vuole risolvere a differenziare i due procedimenti.

In particolare, come detto, Stone propone la matrice diagonale $V = \text{diag}(p)$, mentre Denton propone, nella sua variante PFD (*Proportional First Differences*), la matrice $M = P^{-1}D'DP^{-1}$, dove D è la matrice di differenziazione.

La scelta, nella procedura univariata di Denton, della variante PFD permette di preservare i tassi di crescita ricorrendo alle differenze prime e, contemporaneamente, di modificare le serie in maniera proporzionale al valore assunto dai singoli valori.

Successivamente si è sviluppato il problema della riconciliazione, proponendo come soluzione l'estensione multivariata della procedura di Denton, formulata da Di Fonzo (2002). Il vettore dei dati riconciliati è dato da

$$y = p + VH_r'V_r^{-1}(w - H_r p),$$

dove H_r , V_r e w rappresentano le matrici di rango pieno equivalenti a H , V e z , rispettivamente. La riconciliazione si ottiene quindi aggiungendo ai dati una combinazione lineare delle discrepanze. Per ridurre, in termini computazionali, il problema, Di Fonzo e Marini (2007a) hanno proposto un approccio a due passi della procedura: nel primo le serie vengono sottoposte a benchmarking univariato con la procedura di Denton, poi vengono riconciliati $v < n$ anni simultaneamente (solitamente $v = 1$).

L'approccio di Quenneville e Rancourt (2005) è, concettualmente, un approccio a due passi: dopo aver preservato la dinamica temporale con la procedura univariata di Denton si riconcilia la serie ricorrendo alla metodologia di Stone.

Le tecniche illustrate sono poi applicate alle serie storiche destagionalizzate prodotte con l'indagine mensile sul commercio al dettaglio in Canada, disponibili nel periodo gennaio 1991 - dicembre 2003. I dati sono disponibili secondo due diverse classificazioni: per regioni (province e territori) e per settori commerciali. Il primo lo indicheremo come sistema PR, il secondo come sistema TG. In tutte le procedure abbiamo considerato le somme annuali dei dati grezzi come benchmark temporale.

Inizialmente si è trattato il vincolo come esogeno, ottenuto per somma contemporanea degli aggregati per settore commerciale. Con questo vincolo, suggerito da Statistics Canada, si sono poi riconciliati entrambi i sistemi. Per quel che riguarda il sistema PR, questo presenta discrepanze temporali molto esigue e discrepanze contemporanee che variano, in genere, tra il -1% e l'1%. Le piccole discrepanze temporali comportano che il primo step non modifichi, di fatto, i dati. I risultati che vale la pena segnalare sono:

- l'estensione multivariata della procedura di Denton (DENTF) fornisce globalmente i risultati migliori, ma si caratterizza per RMSA (indice sintetico dell'aggiustamento apportato a ciascuna serie) proporzionali alle dimensioni medie dell'aggregato.
- La procedura di Denton applicata simultaneamente un anno alla volta (DENT1) produce risultati lievemente peggiori rispetto a DENTF.
- La procedura di Quenneville e Rancourt (QR), invece, ottiene risultati mediamente peggiori ma modifica i tassi di crescita in maniera simile tra una serie e l'altra. Questo è confermato da un indice ponderato, che sintetizza gli aggiustamenti apportati dalle procedure di riconciliazione ai tassi di crescita, che fa registrare i risultati migliori proprio per la procedura di Quenneville e Rancourt.

Per evidenziare il ruolo della procedura a due step si è successivamente usato un vincolo esogeno, ottenuto tramite simulazione, nel quale si sono introdotte discrepanze temporali attorno al 5%. Questo ci ha consentito di trarre tre ulteriori considerazioni:

- la procedura DENTF, riconciliando i dati simultaneamente, non è influenzata dalla presenza del primo step e i risultati continuano ad essere molto simili;
- la procedura DENT1, al contrario, beneficia del fatto che la dinamica temporale sia trattata preliminarmente con Denton univariato ed elimina delle differenze che altrimenti si creano tra l'ultimo periodo di un anno e il primo di quello successivo;
- le differenze più evidenti si hanno per la procedura QR che, in questa applicazione, si vede essere molto diversa dalla procedura di Stone.

Una seconda applicazione è stata condotta con un vincolo contemporaneo di tipo endogeno: si vogliono riconciliare i dati in modo che la somma degli aggregati per provincia/territorio sia uguale alla somma degli aggregati per settore commerciale.

Questa applicazione è giustificata da due motivazioni:

- in assenza di un vincolo esogeno fornito dagli esperti del settore e la cui scelta esula dagli obiettivi di questa tesi, un vincolo di questo tipo può sembrare, anche a livello intuitivo, la scelta più naturale;
- con questa procedura valutiamo le potenzialità di applicazione della tecnica, che si mostra può essere estesa al caso in cui la relazione tra aggregati e vincolo non è semplicemente quella di somma. Se nel caso esogeno, per ciascun sistema, la relazione tra gli aggregati e il vincolo è infatti di somma, in questo caso la forma della matrice di aggregazione² è diversa:

$$[1'_{19} \otimes I_{156} \quad -(1'_{12} \otimes I_{156})].$$

Il vincolo contemporaneo implicitamente ricavato dalla procedura di riconciliazione basata su Denton preserva il segno dei tassi di crescita meglio di quanto faccia quello della procedura di Quenneville e Rancourt, con riferimento ai tassi di crescita del vincolo esogeno suggerito dagli esperti di Statistics Canada che teniamo come riferimento.

² In effetti, nell'applicazione proposta nella tesi, la matrice di aggregazione è leggermente diversa, perché deve tenere conto del fatto che il territorio del Nunavut è stato creato nel 1999 e la serie corrispondente non ha valori per i primi 96 periodi (8 anni).

Un possibile spunto di riflessione per eventuali approfondimenti futuri è rappresentato dal seguente quesito: la struttura non casuale delle discrepanze contemporanee può essere utilizzata per migliorare gli esiti dell'aggiustamento, con una procedura di riconciliazione che ne tenga conto?

Infine un accenno a proposito dell'estensione al caso di una riconciliazione a due vie: nella pratica si ha spesso a che fare con serie storiche raccolte secondo due diverse classificazioni. Ci siamo limitati a considerare in questa tesi la riconciliazione a una via, sia per semplificare l'esposizione che per convenienza tipografica. In effetti, le procedure qui descritte possono essere estese anche al caso in cui si considerino contemporaneamente due diversi criteri di aggregazione. Per quanto riguarda la parte applicativa, invece, le difficoltà maggiori sono relative ai tempi computazionali; gli algoritmi di gestione delle matrici sparse presenti in Matlab® e già usati in questa tesi dovrebbero garantire comunque risultati soddisfacenti. Il lettore interessato trova l'applicazione ai dati MRTS in cui si considerano contemporaneamente sia le aggregazioni per provincia/territorio che per settore commerciale in Di Fonzo e Marini (2007a).

Appendice A. Presentazione dei dati

A.1 Struttura di una matrice di contabilità sociale

Totali di riga										
Resto del mondo	Movimenti di reddito									
	Consumo									
	Export									
Risparmio										
Consumo										
Settori istituzionali										
Valore aggiunto	Imposte									
	Fattori produttivi									
Branche d'attività										
Branche d'attività										

Tabella A.1 Struttura di una matrice di contabilità sociale.

A.2 Struttura dei dati

Anno T	Trimestre k	$p_{1,t}$	$y_{01,T}$...	$p_{j,t}$	$y_{0j,T}$...	$p_{M,t}$	$y_{0M,T}$	z_t
1	1	$p_{1,1}$...	$p_{j,1}$...	$p_{m,1}$		zC_1
	2	$p_{1,2}$...	$p_{j,2}$...	$p_{m,2}$		zC_2
	3	$p_{1,3}$...	$p_{j,3}$...	$p_{m,3}$		zC_3
	4	$p_{1,4}$...	$p_{j,4}$...	$p_{m,4}$		zC_4
			$zt_{1,1}$...		$zt_{j,1}$...		$zt_{m,1}$	

	1	$p_{1,4T-3}$...	$p_{j,4T-3}$...	$p_{m,4T-3}$		zC_{4T-3}
	2	$p_{1,4T-2}$...	$p_{j,4T-2}$...	$p_{m,4T-2}$		zC_{4T-2}
	3	$p_{1,4T-1}$...	$p_{j,4T-1}$...	$p_{m,4T-1}$		zC_{4T-1}
	4	$p_{1,4T}$...	$p_{j,4T}$...	$p_{m,4T}$		zC_{4T}
			$zt_{1,T}$...		$zt_{j,T}$...		$zt_{m,T}$	

T	1	$p_{1,4N-3}$...	$p_{j,4N-3}$...	$p_{m,4N-3}$		zC_{4N-3}
	2	$p_{1,4N-2}$...	$p_{j,4N-2}$...	$p_{m,4N-2}$		zC_{4N-2}
	3	$p_{1,4N-1}$...	$p_{j,4N-1}$...	$p_{m,4N-1}$		zC_{4N-1}
	4	$p_{1,4N}$...	$p_{j,4N}$...	$p_{m,4N}$		zC_{4N}
			$zt_{1,N}$			$zt_{j,N}$			$zt_{m,N}$	
	1	$p_{1,n-1}$...	$p_{j,n-1}$...	$p_{m,n-1}$		zC_{n-1}
	2	$p_{1,n}$...	$p_{j,n}$...	$p_{m,n}$		zC_n
N	1	$p_{1,n-1}$...	$p_{j,n-1}$...	$p_{m,n-1}$		zC_{n-1}
	2	$p_{1,n}$...	$p_{j,n}$...	$p_{m,n}$		zC_n

Tabella A.2 Struttura dei dati.

Appendice B. Approfondimenti

B.1 Utilizzo dell'aggiustamento proporzionato Plus-Minus

Si vogliono bilanciare, mediante uso del PMPA, questi valori affinché la loro somma sia pari a 20.

i	p_i	α_i	$y_i = \alpha_i p_i$
1	30	1,044	31
2	-50	0,957	-48
3	35	1,044	37
Totale	15		20

Tabella B.1 Esempio di bilanciamento con il PMPA.

I coefficienti si calcolano immediatamente dalle rispettive formule:

$$\alpha^+ = 1 + \frac{(z - \sum_{i=1}^n p_i)}{\sum_{i=1}^n |p_i|} = 1 + \frac{20 - 15}{115} = 1,044$$

$$\alpha^- = 1 - \frac{(z - \sum_{i=1}^n p_i)}{\sum_{i=1}^n |p_i|} = 1 - \frac{20 - 15}{115} = 0,957.$$

B.2 Algoritmo per lo scaling DSS

Si presenta ora l'algoritmo per la procedura DSS.

(5) *Input*. Una matrice quadrata non negativa P .

(6) *Passo zero (Inizializzazione)*. Si ponga $k = 0$ e $Y^0 = P$.

(7) *Primo passo (Selezione dell'indice)*. Per $i = 1, \dots, n$ si definiscano

$$u_i = \sum_{j=1}^n y_{ij}^k$$

$$v_i = \sum_{j=1}^n y_{ji}^k.$$

Definiamo p come l'indice che soddisfa

$$|u_p - v_p| = \max_{1 \leq i \leq n} |u_i - v_i|.$$

(8) *Secondo passo (Scaling)*. Si definisca α_k tale che

$$\alpha_k u_p = \frac{1}{\alpha_k} v_p,$$

ovvero

$$\alpha_k = \sqrt{\frac{v_p}{u_p}}.$$

(9) *Terzo passo (Aggiornamento)*. Si definisca Y^{k+1} come

$$\begin{cases} \alpha_k y_{ij}^k, & \text{se } i = p, j \neq p \\ \frac{1}{\alpha_k} y_{ij}^k, & \text{se } j = p, i \neq p \\ y_{ij}^k, & \text{altrimenti} \end{cases}$$

(10) *Quarto passo*. Si sostituisca $k \leftarrow k + 1$ e si riparta dal primo passo.

B.3 Il vincolo di aggregazione temporale e dati contraddittori

Quando si parla di vincoli stringenti, si vuole che i totali di ciascuna riga (che rappresentano i diversi aggregati a un certo istante temporale) siano uguali ai totali di ciascuna colonna. Questo implica che i due totali siano tra loro uguali.

Quando questa condizione non è verificata, si parla di dati contraddittori [Bee Dagum e Cholette (2006)]. Questo è quello che accade nel caso analizzato: le somme annuali del totale per l'intera economia canadese, considerate il benchmark contemporaneo, non sono uguali al benchmark contemporaneo.

Formalmente il vincolo che non è rispettato si può scrivere come

$$\sum_{j=1}^m zt_{k,j} \neq \sum_{i=12k-11}^{12k} zc_i.$$

Poiché lo scopo dell'applicazione è riconciliare una tavola di dati destagionalizzati e poiché la stagionalità, per definizione, si esaurisce nell'anno, possiamo considerare i vincoli contemporanei come esatti.

Affinché l'ipotesi di partenza sia soddisfatta, è sufficiente applicare alla serie dei dati zc la procedura di benchmarking univariato PFD di Denton, usando i totali annuali come benchmark.

Al termine di questa semplice operazione si ottiene quindi una nuova serie di dati mensili per l'intera economia canadese che rispetta i vincoli richiesti dalla procedura di bilanciamento.

B.4 Uso dei numeri pseudo casuali nella generazione di vincoli temporali

Lo scopo dell'applicazione descritta nel paragrafo 5.7 è quello di studiare il comportamento delle diverse procedure di riconciliazione in presenza di discrepanze temporali non trascurabili.

Per questo si sono modificati i totali annuali in modo tale da ottenere discrepanze medie del 5% (Figura 5.17). Per farlo, si sono trovati i totali annuali dei dati mensili e si è poi provveduto ad aumentarli di una quantità casuale, i da una variabile casuale uniforme di media 0,05.

```
cas = (rand(169,1));  
cas = cas./10 + 1;
```

Il primo comando genera un vettore di 169 realizzazioni di una variabile casuale di media 0,5; il secondo comando modifica il primo vettore affinché i dati abbiano media 1,05. La moltiplicazione elemento per elemento di questo vettore con il vettore dei totali annuali permette di ottenere le discrepanze temporali desiderate.

B.5 La funzione obiettivo e le matrici di aggregazione in presenza di serie di lunghezza diversa.

Quanto segue è tratto da: Di Fonzo (2009), Nota metodologica per lo script Matlab® ad hoc per la riconciliazione dei dati MRTS small.

Il primo passo consiste nello scrivere gli elementi delle matrici dei dati preliminari, dei dati riconciliati e dei vincoli in forma vettoriale e senza elementi nulli.

$$Y = \begin{bmatrix} y_1 & y_2 & \cdots & 0 \\ 156 \times 1 & 156 \times 1 & & 96 \times 1 \\ & & & y_{13} \\ & & & 60 \times 1 \end{bmatrix} \Rightarrow y = [y'_1 \quad y'_2 \quad \cdots \quad y'_{31} \quad y'_{32}]'_{4896 \times 1}$$

$$P = \begin{bmatrix} p_1 & p_2 & \cdots & 0 \\ 156 \times 1 & 156 \times 1 & & 96 \times 1 \\ & & & p_{13} \\ & & & 60 \times 1 \end{bmatrix} \Rightarrow p = [p'_1 \quad p'_2 \quad \cdots \quad p'_{31} \quad p'_{32}]'_{4896 \times 1}$$

$$ZT = \begin{bmatrix} zt_1 & zt_2 & \cdots & 0 \\ 13 \times 1 & 13 \times 1 & & 8 \times 1 \\ & & & zt_{13} \\ & & & 5 \times 1 \end{bmatrix} \Rightarrow zt = [zt'_1 \quad zt'_2 \quad \cdots \quad zt'_{31} \quad zt'_{32}]'_{408 \times 1}$$

I vincoli di aggregazione, rispettivamente, contemporanea e temporale, possono essere scritti in notazione matriciale nel modo seguente:

Vincolo di aggregazione contemporanea

$$\begin{bmatrix} 1'_{19} \otimes I_{156} & -(1'_{12} \otimes I_{156}) & 0_{96,60} \\ & & -I_{60} \end{bmatrix} y = 0_{156,1} \Leftrightarrow H_C y = 0_{156,1}$$

con

$$H_C = \begin{bmatrix} 1'_{19} \otimes I_{156} & -(1'_{12} \otimes I_{156}) & 0_{96,60} \\ & & -I_{60} \end{bmatrix}.$$

Vincolo di aggregazione temporale

$$\begin{aligned} C_{13 \times 156} &= (I_{13} \otimes 1'_{12}) \\ C_1 &= (I_5 \otimes 1'_{12}) \\ 5 \times 60 \end{aligned} \Rightarrow \begin{aligned} Cy_j &= zt_j \quad j = 1, \dots, 31 \\ C_1 y_{32} &= zt_{32} \end{aligned} \Rightarrow$$

$$\Rightarrow \begin{bmatrix} I_{31} \otimes C & 0_{403,60} \\ 0_{5,4836} & C_1 \end{bmatrix} y = zt \Leftrightarrow H_T y = zt$$

con

$$H_T = \begin{bmatrix} I_{31} \otimes C & 0_{403,60} \\ 0_{5,4836} & C_1 \end{bmatrix}.$$

Indicati con H la matrice $H = \begin{bmatrix} H_C \\ H_T \end{bmatrix}$ e con z il vettore $z = [0_{1,156} \quad zt']$, il complesso dei vincoli (ridondanti) è quindi dato da

$$Hy = z.$$

Assumendo che nessun valore preliminare sia nullo, la funzione obiettivo da minimizzare è

$$\sum_{j=1}^{31} \sum_{t=2}^{156} \left(\frac{y_{j,t}}{p_{j,t}} - \frac{y_{j,t-1}}{p_{j,t-1}} \right)^2 + \sum_{t=2}^{60} \left(\frac{y_{32,t}}{p_{32,t}} - \frac{y_{32,t-1}}{p_{32,t-1}} \right)^2 \quad \text{Proportional First Differences (PFD), variante esatta}$$

Per scriverla in notazione matriciale, si considerino le matrici

$$\Omega = \begin{bmatrix} I_{31} \otimes \Delta' \Delta & 0_{4836,60} \\ 0_{60,4836} & \Delta'_1 \Delta_1 \end{bmatrix} \text{ e } \Omega_{PFD} = \hat{p}^{-1} \Omega \hat{p}^{-1},$$

dove $\hat{p} = \text{diag}(p)$, mentre Δ è la matrice (155 x 156) delle differenze prime operanti su ciascuna delle prime dodici serie e Δ_1 è la matrice (59 x 60) delle differenze prime operanti sulla tredicesima serie, come già definite nella (3.6).

La soluzione secondo la procedura di Denton PFD simultaneo si ottiene dunque risolvendo il problema di minimizzazione quadratica vincolata (PFD). Applicando il metodo dei moltiplicatori di Lagrange, il sistema da risolvere è dato da

$$\begin{bmatrix} \Omega_{PFD} & H' \\ H & 0_{564 \times 564} \end{bmatrix} \begin{bmatrix} y \\ \lambda \end{bmatrix} = \begin{bmatrix} \Omega_{PFD} p \\ z \end{bmatrix}$$

in cui λ è un vettore di moltiplicatori di Lagrange. Come già dettagliatamente illustrato nel capitolo 4, si ricava che i valori riconciliati si ottengono risolvendo il sistema:

$$\begin{bmatrix} \Omega & \hat{p}H' \\ H\hat{p} & 0_{564 \times 564} \end{bmatrix} \begin{bmatrix} \hat{p}^{-1}y \\ \lambda \end{bmatrix} = \begin{bmatrix} 0_{4896,1} \\ z \end{bmatrix} \Leftrightarrow Ax = b$$

in cui $x = \begin{bmatrix} \hat{p}^{-1}y \\ \lambda \end{bmatrix}$ e $b = \begin{bmatrix} 0_{4896,1} \\ z \end{bmatrix}$ sono vettori di quantità note, e $A =$

$\begin{bmatrix} \Omega & \hat{p}H' \\ H\hat{p} & 0_{564 \times 564} \end{bmatrix}$ è la matrice (singolare) dei coefficienti del sistema lineare.

Indicato quindi con x^* il vettore contenente i primi 4896 valori del vettore soluzione x , il vettore contenente le serie riconciliate ($y \equiv x_{sol}$) si ottiene come $y = \hat{p}x^*$.

Appendice C. Alcuni script di esempio

C.1 Bilanciamento di una matrice con il metodo Stone

L'esempio qui proposto, tratto da Marini (2007), propone lo svolgimento di un problema di bilanciamento su una matrice di dimensioni molto piccole, per capire la soluzione senza soffermarsi sui problemi computazionali che emergono quando si lavora con dati reali, che coinvolgono matrici di dimensioni notevoli.

Immaginiamo un sistema economico con tre imprese (A, B, C) che utilizzano tre prodotti (1, 2, 3) nel loro processo produttivo. La seguente tabella riassume le informazioni preliminari disponibili:

	A	B	C	Totale
P1	131.25	168.75	75.00	400.00
P2	200.00	405.00	180.00	850.00
P3	78.75	101.25	45.00	250.00
Totale	500.00	700.00	350.00	

Tabella C.1 Matrice dei dati preliminari.

Si noti come questi dati non abbiamo un unico totale, ma tre totali differenti: la somma delle celle interne (quella matrice che abbiamo chiamato matrice P) ha totale 1350, le marginali di riga (z_r) sommano a 1500 e le marginali di colonna (z_c) a 1550.

I coefficienti di alterabilità associati a ciascun valore della tabella sono:

	A	B	C	Totale
P1	0.50	0.50	0.50	0.10
P2	0.01	0.50	0.50	0.50
P3	0.50	0.50	0.50	1.00
Totale	0.50	0.10	0.50	

Tabella C.2 Matrice dei coefficienti di alterabilità.

Il vettore di stime preliminari sarà composto dai 15 elementi della matrice dei costi, vettorializzati nella maniera seguente:

$$p_z = \begin{bmatrix} P_{11} \\ P_{21} \\ P_{31} \\ P_{12} \\ P_{22} \\ P_{32} \\ P_{13} \\ P_{23} \\ P_{33} \\ z_r \\ z_c \end{bmatrix} = \begin{bmatrix} 131.25 \\ 168.75 \\ 75.00 \\ 200.00 \\ 405.00 \\ 180.00 \\ 78.75 \\ 101.25 \\ 45.00 \\ 400.00 \\ 850.00 \\ 250.00 \\ 500.00 \\ 700.00 \\ 350.00 \end{bmatrix}.$$

La matrice V è composta dai seguenti elementi (si riportano solo i termini sulla diagonale, essendo tutti gli altri nulli):

$$diag(V) = \begin{bmatrix} 0.50 \\ 0.50 \\ 0.50 \\ 0.01 \\ 0.50 \\ 0.50 \\ 0.50 \\ 0.50 \\ 0.50 \\ 0.50 \\ 0.10 \\ 0.50 \\ 1.00 \\ 0.50 \\ 0.10 \\ 0.50 \end{bmatrix} p_z = \begin{bmatrix} 65.6 \\ 84.4 \\ 37.5 \\ 2.0 \\ 202.5 \\ 90.0 \\ 39.4 \\ 50.6 \\ 22.5 \\ 40.0 \\ 425.0 \\ 250.0 \\ 250.0 \\ 70.0 \\ 175.0 \end{bmatrix}$$

I calcoli sono stati svolti, per rapidità, con l'ambiente per il calcolo numerico e linguaggio di programmazione Matlab®.

```
p = [131.25 168.75 75 200 405 180 78.75 101.25 45]
zr = [400 850 250]
zc = [500 700 350]
pz = [p zr zc]
pz = pz'
```

```

a = [0.5 0.5 0.5 0.01 0.5 0.5 0.5 0.5 0.5 0.1 0.5 1 0.5 0.1 0.5]
a=a'

V = diag(a*pz')
V = diag(V)

G11 = [1 1 1 0 0 0 0 0 0; 0 0 0 1 1 1 0 0 0 ; 0 0 0 0 0 0 1 1 1]
G12 = [eye(3) eye(3) eye(3)]
G2 = eye(6)*(-1)
G1 = [G11;G12]
G = [G1 G2]
y = pz - V*G' * (G*V*G')^(-1) * G*pz

```

La matrice bilanciata è data da:

	A	B	C	Totale
P1	148.0	171.4	80.3	399.7
P2	200.6	423.1	197.9	821.6
P3	89.7	103.9	48.7	242.3
Totale	438.3	698.4	326.9	1463.6

Tabella C.3 Matrice dei dati bilanciati.

Le seguenti tabelle mostrano le differenze tra i dati iniziali e i dati bilanciati

	A	B	C	Totale
P1	16.8	2.6	5.3	-0.3
P2	0.6	18.1	17.9	-28.4
P3	10.9	2.6	3.7	-7.7
Totale	-61.7	-1.6	-23.1	

Tabella C.4 Matrice delle discrepanze (in valore assoluto).

e le differenze percentuali

	A	B	C	Totale
P1	11.3%	1.5%	6.6%	-0.1%
P2	0.3%	4.3%	9.0%	-3.5%
P3	12.2%	2.6%	7.5%	-3.2%
Totale	-14.1%	-0.2%	-7.1%	

Tabella C.5 Matrice delle discrepanze (percentuali).

Si noti come i valori a cui era stata assegnata una grande affidabilità sono rimasti pressoché immutati. Il totale complessivo risulta pari a 1463.6, che è una combinazione lineare dei tre totali di partenza.

C.2 Aggiustamento secondo la procedura univariata di Denton

Per acquisire pratica con il calcolo matriciale e familiarizzare con questi concetti chiave si è ripetuto l'esempio proposto da Denton.

La serie originale presenta dei dati trimestrali ($\tau = 4$) che coprono un periodo di tempo di $N = 5$ anni, per un totale di $n = 20$ osservazioni. Questi dati si ripetono uguali ogni anno e la loro somma annuale è 400. I totali annuali ai quali la serie deve essere aggiustata sono 500, 400, 300, 400 e 500.

Per facilitare i calcoli è stato realizzato un semplice script per il linguaggio open source R:

```
aggiusta<-function(p,y0,h=0,prop="F",forw="F") {
  # Matrice identità
  id<-function(n){
    I<-matrix(0,n,n); diag(I)<-1
    I
  }

  # Matrice di aggregazione temporale
  at<-function(n,N){
    J<-matrix(0,N,n)
    k<-n/N
    for (R in 1:N) {a<-k*(R-1)+1; b<-k*R; J[R,a:b]<-1}
    J
  }

  # Matrice differenze h-esime
  md<-function(n,h=1, forw="F"){
    D1<-id(n)
    for (R in 2:n) {D1[R,R-1]<-(-1)}
    if (forw=="T") D1<-(-1)*t(D1)
    D<-D1
    if (h>=2) {
      for (H in 2:h) {D<-D1%*%D}
    }
    D
  }

  if(h==0) M<-id(length(p))
  if(h>=1) {
    D<-md(length(p),h,forw)
    M<-t(D)%*%D
  }

  J<-at(length(p),length(y0))
  tJ<-t(J)

  if (prop=="T") {
    Z<-id(length(p));
    diag(Z)<-p;
    M<-solve(Z)%*%M%*%solve(Z)
  }
}
```



```

    }

    iM<-solve(M)
    C<-iM%*%tJ%*%solve(J%*%iM%*%tJ)
    r<-y0-J%*%p
    y<-p+C%*%r

    as.vector(y)
  }

```

La tabella riporta i valori delle serie aggiustate:

		Serie	Serie aggiustate secondo le diverse funzioni di perdita							
Anno e trimestre		originaria	naive	AFD	ASD	ATD	pro-rata	PFD	PSD	PTD
		p	$(x - z)$	$\Delta(x - z)$	$\Delta^2(x - z)$	$\Delta^3(x - z)$	$(\frac{x - z}{z})$	$\Delta(\frac{x - z}{z})$	$\Delta^2(\frac{x - z}{z})$	$\Delta^3(\frac{x - z}{z})$
1	1T	50	75	67	62	59	56	57	55	54
	2T	100	125	127	125	123	122	124	122	120
	3T	150	175	180	182	184	200	194	194	195
	4T	100	125	126	130	134	122	125	129	132
2	1T	50	50	65	70	74	50	58	61	62
	2T	100	100	105	106	107	100	107	109	111
	3T	150	150	145	142	141	150	146	145	144
	4T	100	100	85	81	78	100	89	85	83
3	1T	50	25	27	24	22	44	40	39	38
	2T	100	75	73	72	71	78	74	73	72
	3T	150	125	123	124	125	100	109	110	111
	4T	100	75	78	80	81	78	77	78	79
4	1T	50	50	37	38	39	50	53	43	43
	2T	100	100	96	96	96	100	94	94	94
	3T	150	150	154	155	154	150	153	154	153
	4T	100	100	112	112	111	100	110	110	110
5	1T	50	75	69	68	67	56	58	58	58
	2T	100	125	124	123	123	122	123	121	122
	3T	150	175	178	178	178	200	190	189	190
	4T	100	125	129	132	132	122	129	131	130

Tabella C.6 Valori originari e valori aggiustati secondo la procedura di Denton e diverse funzioni di perdita.

C.3 Il modello di regressione per la riconciliazione

Quenneville e Rancourt (2005) propongono una semplice implementazione della loro idea in SAS®. Lo riproponiamo usando la funzione `lse` di MATLAB®, scritta da John D'Errico per la soluzione di problemi di minimizzazione quadratica vincolata.

Questi sono i dati che utilizziamo: il vettore p dei dati preliminari, la matrice X e la matrice G . La somma dei valori aggiustati deve essere pari a z .

Nell'ottica del modello di regressione, possiamo scrivere

$$p = Xy \quad \text{soggetta al vincolo } Gp = z$$

$$p = \begin{bmatrix} 5 \\ 25 \end{bmatrix} \quad X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad G = \begin{bmatrix} 1 & 1 \end{bmatrix} \quad z = 40$$

La funzione `lse` minimizza la differenza tra $Xy - p$ rispettando il vincolo imposto e accetta in input, rispettivamente, la matrice delle variabili dipendenti, X , quella delle variabili indipendenti, p , la matrice che esprime il tipo di vincolo, G , e il valore del vincolo z . Infine, il parametro `'\'` indica la metodologia (è il default), mentre l'ultimo è il vettore dei pesi, inversamente proporzionali alla varianza (che è supposta pari alle stime preliminari dei dati).

```
X = [1 0; 0 1];
p = [5 25]';
G = ones(1,2);
z = 40;
lse(X,p,G,z,'\ ',1./p)

ans =

    6.6667
   33.3333

alter = [0.1 1 1]';
X = [1 1; 1 0; 0 1];
p = [40 5 25]';
y = lse(X,p,[],[],'\ ',(1./(alter.*p)))

y =

    6.4706
   32.3529
```

```
y = lse(X,p,[],[],'\',(1./(alter.*p)))
y =
    6.6666
   33.3332
```

Un software statistico permette anche ad un utente di livello medio-basso di effettuare la regressione, perché può usare i pochi comandi che già conosce (se non addirittura un'interfaccia grafica) senza addentrarsi nella costruzione di matrici. Ancora una volta la scelta è ricaduta su R, molto diffuso in ambito universitario e con enormi potenzialità di crescita legate sia al tipo di licenza, open source, sia alla sempre maggiore disponibilità di interfacce grafiche che lo portano a essere preferito, in molti ambienti dove si usa la statistica senza specifiche competenze, al più noto SPSS.

Purtroppo la funzione `lm` di R, usata per stimare modelli di regressione lineare, non permette di aggiungere dei vincoli al modello. Il programma può essere usato anche nel caso di vincoli stringenti tramite il comando `solve.QP`, nel pacchetto `quadprog`. Questo però richiede un'impostazione di tipo matriciale, la stessa richiesta in Matlab®, e contraddice lo scopo descritto al paragrafo precedente. Proponiamo quindi solo lo script per risolvere il problema del bilanciamento nel caso di vincoli non stringenti tramite l'uso del comando `lm`. Il lettore interessato non troverà particolari difficoltà a riscrivere per R il codice Matlab® già riportato.

Sia P la matrice dei dati da aggiustare:

5	8	14
7	4	10
12	12	22

Tabella C.7 Matrice dei dati preliminari.

e siano i coefficienti di alterabilità pari a 1 per i valori interni e pari a 0,1 per i totali.

Il codice R per ottenere la stima tramite regressione è:

```
#Imputazione dei dati
p <- matrix(c(5,8,14,7,4,10,12,12,22),ncol=3,byrow=T)
```

```
a <- matrix(c(1,1,0.1,1,1,0.1,0.1,0.1,0.1),ncol=3,byrow=T)
w <- 1/(a*p)
```

```
#Creazione della matrice X
creax<-function(p){
  n <- dim(p)[1]-1
  m <- dim(p)[2]-1
  a1 <- diag(1,n)
  b1 <- matrix(c(diag(1,m),rep(1,m)),ncol=m,byrow=T)
  x1 <- kronecker(a1,b1)
  a2 <- matrix(c(1,1),ncol=n)
  b2 <- diag(1,m)
  x2 <- kronecker(a2,b2)
  x <- matrix(-1,nrow=nrow(x1)+n+1,ncol=ncol(x1))
  x[1:(nrow(x1)),] <- x1
  x[(nrow(x1)+1):(nrow(x1)+n),] <- x2
  x[(nrow(x1)+n+1),] <- 1
  x
}
```

```
aggiusta <- function(p,x=creax(p),a,w=(1/(c(p)*c(a)))) {
  st <- lm(c(p)~x-1,weights=w)$coef
  y0 <- matrix(st,ncol=(ncol(p)-1))
  y <- matrix(-1,nrow=nrow(y0)+1,ncol=ncol(y0)+1)
  y[1:nrow(y0),1:ncol(y0)]<-y0
  y[nrow(y0)+1,1:ncol(y0)]<-apply(y0,2,sum)
  y[1:nrow(y0),ncol(y0)+1]<-apply(y0,1,sum)
  y[nrow(y0)+1,ncol(y0)+1]<-sum(y0)
  y
}
```

```
aggiusta(p,a=a)
```

Il numero di righe di codice può far credere che il procedimento sia meno semplice di quanto descritto. In pratica invece, una volta inseriti i dati e creata la matrice delle X , cioè la matrice che spiega le relazioni causali esistenti nel modello di regressione, il comando è quello evidenziato in grassetto. Qui è stato inserito all'interno di una funzione che restituisce i dati in forma tabellare, ipotizzando che anche l'input sia inserito nella stessa maniera. Nell'esempio sopra riportato il risultato è

```
      [,1]      [,2]      [,3]
[1,]  5.304502  8.202398 13.50690
[2,]  6.329802  3.474627  9.80443
[3,] 11.634305 11.677026 23.31133
```

La scelta di coefficienti di alterabilità molto piccoli permette di ottenere vincoli che si possono considerare stringenti.

C.4 Funzione Matlab® per il benchmarking univariato secondo la procedura di Denton

Questa funzione è una modifica della funzione `udenton` proposta da Di Fonzo e Marini. La parte centrale della funzione, evidenziata in grassetto, è dovuta a loro.

```
function [ybench,et] = udbp (prel, y0, bench, tb, aggr)
%   UDBP
%   Univariate Denton's benchmarking procedure
%
%   -----
%   [ybench,et] = udbp (prel, y0, bench, tb, aggr);
%   -----
%
% INPUT:
%   prel      : preliminary series to be benchmarked (sn x m)
%   y0        : vector of temporally aggregated constraints (n x m)
%   bench     : benchmarking estimation
%               1: Denton additive first differences
%               2: Denton additive second differences
%               3: Denton proportional first differences
%               4: Denton proportional second differences
%   tb        : 1: Approximate first differences matrix
%               2: Exact first differences matrix
%   aggr      : type of temporal aggregation
%               1: sum
%               2: average
%               3: sample of the last
%               4: sample of the first
%
% OUTPUT:
%   ybench    : vector of benchmarked series (sn x 1)
%   et        : elapsed time
%
% REFERENCES:
%   Denton F.T. (1971), 'Adjustment of monthly or quarterly series
to annual totals:
%   An approach based on quadratic minimization', Journal of the
%   American Statistical Association, 66: 99-102.
%
% Version number: 0.3
% Last update: January 13, 2008
% =====*/

% Initialize output arguments
ybench = [];
et = 0;

% Setting of the time counter
tic;

[n,m] = size(y0); % number of low-frequency observations
and variables
[sn,m_prel] = size(prel); % number of high-frequency observa-
tions
s = floor(sn/n);
```

```

%-----
% Check on input variables
%
% sn      must be at least >= s*n for the program to work
% bench   must be 1,2,3 or 4
% aggr    must be 1,2,3 or 4
% m       must be equal to m_prel for the program to work
% -----

if (sn < s*n)
    disp('UDBP - INPUT CHECKS FAILED');
    disp('Too few observations on the preliminary series. ');
    disp('Check your data!');
    disp('UDBP stops');
    return
end;

if (bench~=1) && (bench~=2) && (bench~=3) && (bench~=4)
    disp('UDBP - INPUT CHECKS FAILED');
    disp('The flag for the benchmarking technique is wrong. ');
    disp('Check your data!');
    disp('UDBP stops');
    return;
end;

if (aggr~=1) && (aggr~=2) && (aggr~=3) && (aggr~=4);
    disp('UDBP - INPUT CHECKS FAILED');
    disp('Error in the aggregation type. ');
    disp('It must be 1, 2, 3 or 4. Check your program!');
    disp('UDBP stops')
    return;
end;

if (m~=m_prel);
    disp('UDBP - INPUT CHECKS FAILED');
    disp('Number of preliminary series must be equal to number of ag-
gregated series. ');
    disp('Check your data!');
    disp('UDBP stops')
    return;
end;

% End of input checks
% -----

for i=1:m
    prel0 = prel(:,i);
    y00 = y0(:,i);

% Check if all data in a low-frequency period are zeros and remove
alpha_hf = [];
alpha_lf = [];
for count = 1:n
    start = s*(count-1) + 1;
    stop = count*s;
    if (sum(abs(prel0(start:stop,1)))>0)
        alpha_hf = [alpha_hf start:stop];
        alpha_lf = [alpha_lf count];
    end
end

```

```

        end;
    end;
    [m0, alpha_n] = size(alpha_lf);
    if (alpha_n ~= n)
        prel0 = prel0(alpha_hf,1);
        y00 = y00(alpha_lf,1);
        n = alpha_n;
        [sn,m0] = size(prel0);
    end;

    % Forecasts check
    if sn>s*n
        forecast = sn-s*n;
    else
        forecast = 0;
    end;

    % C: temporal aggregation matrix
    switch aggr
        case 1
            C = kron(eye(n),ones(1,s)); % flows
        case 2
            C = kron(eye(n),ones(1,s)/s); % index
        case 3
            C = kron(eye(n),[zeros(1,s-1),1]); % end-of-period
    stock
        case 4
            C = kron(eye(n),[1,zeros(1,s-1)]); % beginning-of-
    period stock
    end;

    if forecast>0
        C = [C,zeros(n,forecast)];
    end;

    % First Differences matrices
    D = eye(sn)-[zeros(1,sn-1);eye(sn-1),zeros(sn,1)]; % approximate
    first differences (sn x sn)
    D1 = D(2:sn,1:sn); % exact first
    differences (sn-1 x sn)
    D2 = D(3:sn,2:sn); % exact second
    differences (sn-2 x sn-1)

    if (tb==1) % Benchmarking using D (approximate first differences
    matrix)
        switch bench
            case 1
                V = inv(D'*D); % Denton AFD
            case 2
                V = inv(D'*D'*D*D); % Denton ASD
            case 3
                preldiag = diag(prel0);
                V = inv(D'*D);
                V = preldiag*V*preldiag; % Denton PFD
            case 4
                preldiag = diag(prel0);
                V = inv(D'*D'*D*D);
                V = preldiag*V*preldiag; % Denton PSD
        end;
    end;

```

```

V0inv = inv(C*V*C');
ybench(alpha_hf,i) = prel0+V*C'*V0inv*(y00-C*prel0);
elseif (tb==2) % Benchmarking using D1 and D2 (exact first differences matrices)
    switch bench
    case 1 % Denton additive
first differences
        Mmat=D1'*D1;
        sysmat = [[Mmat,C'] ; [C,zeros(n,n)]];
        syssol=inv(sysmat)*[Mmat*prel0 ; y00];
    case 2 % Denton additive
second differences
        Mmat=D1'*D2'*D2*D1;
        sysmat = [[Mmat,C'] ; [C,zeros(n,n)]];
        syssol=inv(sysmat)*[Mmat*prel0 ; y00];
    case 3 % Denton proportional
first differences
        prel0 = prel0 ./ kron(mean(prel0)',ones(sn,1)); % scales
the preliminary series
        Mmat=D1'*D1;
        preldiag = diag(prel0);
        preldiaginv = inv(preldiag);
        preldiag1= diag([prel0;ones(n,1)]);
        sysmat = [ [Mmat,(preldiag*C')] ;
        [(C*preldiag),zeros(n,n)] ];
        syssol=preldiag1*inv(sysmat)*[zeros(sn,1) ; y00];
    case 4 % Denton proportional
second differences
        prel0 = prel0 ./ kron(mean(prel0)',ones(sn,1)); % scales
the preliminary series
        Mmat=D1'*D2'*D2*D1;
        preldiag = diag(prel0);
        preldiag1= diag([prel0;ones(n,1)]);
        sysmat = [ [Mmat,(preldiag*C')] ;
        [(C*preldiag),zeros(n,n)] ];
        syssol=preldiag1*inv(sysmat)*[zeros(sn,1) ; y00];
    end;

ybench(alpha_hf,i) = syssol(1:sn,:);

end;
end;

% Reconciled data verification
[n,m] = size(y0); % number of low-frequency observations
and variables
[sn,m_prel] = size(prel); % number of high-frequency observations
s = floor(sn/n);
y0_z = (kron(eye(n),ones(1,s)))*ybench;
TA = round(sum(sum((y0-y0_z))));
if (TA~=0)
    disp('Errore nel benchmarking');
end

% Elapsed time
et = toc;
return

```


Bibliografia

Almon, C. (1968). Recent methodological advances in input-output in the United States and Canada, presentato a "Fourth International Conference on Input-Output Techniques", Genova.

Antonello, P. (1989). Bilanciamento simultaneo dei conti economici nazionali, a prezzi correnti ed a prezzi costanti, *Economia, Società e Istituzioni*, volume 1, n° 2, pagine 347-368.

Antonello, P. (1990). Simultaneous balancing of Input-Output tables at current and constant prices with first order vector autocorrelated errors, *Economic System Research*, volume 2, n° 2, pagine 157-172.

Bartholdy, K. (1987). A new method for balancing the national accounts, Working Paper, 87/66, International Monetary Fund, Washington, D. C.

Bee Dagum, E. (1978). Modelling, Forecasting and Seasonally Adjusting Economic Time Series with the X-11-ARIMA Method, *The Statistician*, volume 27, pagine 203-216.

Bee Dagum, E. e Cholette, P. A. (2006). Benchmarking, temporal distribution, and reconciliation methods for time series, *Lecture Notes in Statistics*, 186, Springer, New York.

Bloem, A. M., Dippelsman, R. J. e Mæhle, N. Ø. (2001). Quarterly National Accounts Manual. Concepts, Data Sources, and Compilation, International Monetary Fund, Washington D.C.

Bregman, L. (1967). Proof of convergence of Sheleikhovskii's method for a problem with transportation constraints, *USSR Computational Mathematics and Mathematical Physics*, volume 1, pagine 191-204.

Byron, R. P. (1978). The estimation of large social account matrices, *Journal of the Royal Statistical Society*, volume 141, n° 3, pagine 359-367.

Cappuccio, N. e Orsi, R. (2005). *Econometria*, Il Mulino, Bologna.

Caramaschi, E. (2004). *Matrice di contabilità sociale e distribuzione del reddito. Un'applicazione al caso italiano*, IRPET, Firenze.

Deming, W. e Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known, *Annals of Mathematical Statistics*, volume n° 11, pagine 427-444.

Denton, F. T. (1971). Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization, *Journal of the American Statistical Association*, volume 66, n° 333, pagine 99-103.

Di Fonzo, T. (1994). Temporal disaggregation of a system of time series when the aggregate is known. Optimal vs adjustment methods, presentato a "INSEE-Eurostat Quarterly National Accounts workshop", Dicembre 2004, Paris-Bercy.

Di Fonzo, T. (2002). Temporal disaggregation of a system of time series when the aggregate is known. Optimal vs adjustment methods, presentato a "INSEE-Eurostat Quarterly National Accounts workshop", Dicembre 1994, Paris-Bercy.

Di Fonzo, T. (2003). Benchmarking di serie storiche economiche. Nota tecnica ed estensioni., Dipartimento di Scienze Statistiche, Working Paper 2003.10, Padova.

Di Fonzo, T. (2008). Notes on reconciliation procedures: RAS, least squares and others, non pubblicato.

Di Fonzo, T. e Marini, M. (2003). Benchmarking systems of seasonally adjusted time series according to Denton's movement preservation principle, Working Paper, 2003.09, Dipartimento di Scienze Statistiche, Padova.

Di Fonzo, T. e Marini, M. (2005). Benchmarking systems of seasonally adjusted time series, *Journal of Business Cycle Measurement and Analysis*, volume 2, pagine 89-123.

Di Fonzo, T. e Marini, M. (2006). Benchmarking a system of time series: Denton's movement preservation principle vs. a data based procedure, Working Paper 2006, no KS-DT-05-008-EN, Eurostat, Luxembourg.

Di Fonzo, T. e Marini, M. (2007a). Reconciliation of systems of time series: methods and applications to economic statistics, non pubblicato.

Di Fonzo, T. e Marini, M. (2007b). Statistical reconciliation of time series. Movement preservation vs a data based procedure, Working Paper, 2007.10, Dipartimento di Scienze Statistiche, Padova.

Di Fonzo, T. e Marini, M. (2007c). Techniques for reconciling systems of time series: an application to the european quarterly sectorial accounts, non pubblicato.

Fofana, I., Lemelin, A. e Cockburn, J. (2005). Balancing a social accounting matrix: theory and application, Modeling and Policy Impact Analysis (MPIA): Training material per il Poverty and economic policy research network, Centre interuniversitaire sur le risque, les politiques économiques et l'emploi; Université Laval, Laval, Quebec (Canada).

Friedlander, D. (1961). A technique for estimating a contingency table, given the marginal totals and some supplementary data, *Journal of the Royal Statistical Society*, volume A, n° 124, pagine 412-420.

Guerrero, V. M. e Nieto, F. H. (1999). Temporal and contemporaneous disaggregation of multiple economic time series, *TEST*, volume 8(2), pagine 459-489.

Harrigan, F. (1990). The reconciliation of inconsistent economic data: the information gain, *Economic Systems Research*, volume 2, n° 1, pagine 17-25.

Harville, D. A. (1997). Matrix algebra from a statistician's perspective, Springer, New York.

Helfand, S. D., Monsour, N. J. e Trager, M. L. (1977). Historical Revision of Current Business Survey Estimates, in Proceedings of the Business and Economic Statistics Section, American Statistical Association, pagine 246-250, Washington.

Higham, D. J. e Higham, N. J. (2005). Matlab guide, Society for Industrial and Applied Mathematics, Filadelfia.

Istat (2001). I conti degli italiani, Il Mulino, Bologna.

Istat (2007a). Annuario statistico italiano 2007, Capitolo XII, pagine 301-323, Istat, Roma.

Istat (2007b). Le tavole delle risorse e degli impieghi e la loro trasformazione in tavole simmetriche. Nota metodologica a "Conti nazionali: il nuovo sistema input output", Istat, Roma.

Jackson, R. W. e Murray, A. T. (2004). Alternative input-output matrix updating formulations, Economic System Research, volume 16, n° 2, pagine 135-148.

Junius, T. e Oosterhaven, J. (2003). The solution of updating or regionalizing a matrix with positive and negative entries, Economic System Research, volume 15, n° 1, pagine 87-96.

Kruithof, J. (1937). Telefoonverkeersrekening, De Ingenieur, volume 52, n° 8, pagine E15-E25.

Lahr, M. L. e De Mesnard, L. (2004). Biproportional techniques in Input-Output analysis: table updating and structural analysis, Economic System Research, volume 16, n° 2, pagine 115-134.

Laniel, N. (1986). Adjustment of economic production sub-annual series to corresponding annual series, Statistics Canada, Ottawa.

Lavanda, I., Rampa, G. e Soro, B. (1999). La revisione delle tavole intersettoriali 1970-90: metodo e procedure, in Rivista di Statistica Ufficiale, volume 1/1999, I-STAT, pagine 23-79, Roma.

Lavanda, I., Rampa, G., Soro, B. e Stella, L. (1996). La stima disaggregata di flussi di produzione interna e di importazione nella contabilità intersettoriale. I - Aspetti metodologici, Quaderni dell'Istituto di Scienze Economiche e Finanziarie, 15, Università degli studi, Genova.

Luenberger, D. G. (2003). Linear and nonlinear programming, Springer, New York.

Magnus, J. R. e Neudecker, H. (1988). Matrix differential calculus with applications in statistics and econometrics, Wiley, New York.

Marini, M. (2007). Il bilanciamento secondo il metodo Stone: formulazione del problema e soluzioni alternative, presentato a "Il bilanciamento di informazioni statistiche con diversa affidabilità. Casi di studio: aspetti teorici e soluzioni pratiche", 29 Ottobre 2007, Contabilità Nazionale, Istat, Roma.

Nicolardi, V. (1998). Un sistema di bilanciamento per matrici contabili di grandi dimensioni, in Quaderni di Ricerca, volume 4/1998, ISTAT, Roma.

Nicolardi, V. (2000). Balancing large accounting systems: an application to the 1992 Italian I-O Table., presentato a "XIII International Conference on Input-Output Techniques", 21-25 Agosto 2000, Università di Macerata, Macerata.

Oosterhaven, J. e Talsma, B. G. (2005). GRAS and RAS versus minimizing absolute and squared differences in coefficients, presentato a "XV International Conference on Input-Output Techniques", 27 Giugno - 1 Luglio 2005, Pechino, Cina.

Pyatt, G. e Round, G. I. (1977). Social accounting matrices for development planning, Review on Income and Wealth, volume 23, n° 4, pagine 339-364.

Quenneville, B. e Rancourt, E. (2005). Simple methods to restore the additivity of a system of time series, presentato a "Workshop on frontiers in benchmarking techniques and their application to official statistics", 8 aprile 2005, OECD, Lussemburgo.

Schneider, M. H. (1989a). Matrix Scaling, Entropy Minimization, and Conjugate Duality (I): Positivity Conditions, *Linear Algebra and its Applications*, volume 114/115, pagine 785-813.

Schneider, M. H. (1989b). Matrix Scaling, Entropy Minimization, and Conjugate Duality (II): The Dual Problem, *Mathematical Programming*, volume 48, n° 1-3, pagine 103-124.

Schneider, M. H. e Zenios, S. A. (1990). A comparative study of algorithms for matrix balancing, *Operations Research*, volume 38, n° 3, pagine 439-455.

Siesto, V. (1996). *La contabilità nazionale italiana*, Il Mulino, Bologna.

Sjöberg, L. (1982). Jämförelse av Uppräkningsmetoder för Nationalräkenskapsdata (Comparison of Adjustment Methods for National Accounts Data), Memorandum, Statistics Sweden, Stockholm.

Skjæveland, A. (1985). Avstemming av Kvartalsvise Nasjonalregnskapsdata mot Årlige Nasjonalregnskap (Reconciliation of Quarterly National Accounts Data Against Annual National Accounts), Interne notater, 85/22, Statistics Norway, Oslo.

Statistics Canada (2008). *Retail Trade*, vol. 80 n° 8, Statistics Canada, Ottawa.

Stone, R. (1961). *Input-Output and national accounts*, OECD, Parigi.

Stone, R. (1962). A social accounting matrix for 1960, Chapman & Hall per "The Department of Applied Economics" dell'Università di Cambridge, Londra.

Stone, R. (1990). Adjusting the national accounts, in *Annali di Statistica*, volume IX, Nuova contabilità nazionale, n° IX, Istat, pagine 311-325, Roma.

Stone, R., Champernowne, D. G. e Meade, J. E. (1942). The precision of national income estimates, *The Review of Economic Studies*, volume 9, n° 2, pagine 111-125.

Theil, H. (1961). *Economic forecasts and policy*, North Holland, Amsterdam.

Van der Ploeg, F. (1982). Reliability and the adjustment of sequences of large economic accounting matrices, *Journal of the Royal Statistical Society. Series A (General)*. volume 142, n° 2, pagine 169-194.

Weale, M. (1988). The reconciliation of values, volumes and prices in the national accounts, *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, volume 151, n° 1, pagine 211-221.